

Kennistoetsing bij huisartsen

Citation for published version (APA):

Pollemans, M. (1994). *Kennistoetsing bij huisartsen*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.19941201mp>

Document status and date:

Published: 01/01/1994

DOI:

[10.26481/dis.19941201mp](https://doi.org/10.26481/dis.19941201mp)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

STELLINGEN

behorend bij het proefschrift

Kennistoetsing bij huisartsen

van

Marjan Pollemans

1. Een kennistoets voor praktizerende huisartsen moet qua inhoud zo nauw mogelijk aansluiten bij de dagelijkse praktijkuitoefening.
2. Om goed inzicht te krijgen in de eigen kennislacunes, is het huisartsen eerder aan te raden af te gaan op door hen behaalde resultaten op kennistoetsen, dan op de eigen inschatting van deze lacunes.
3. Het uitsluitend gebruiken van kennistoetsen om een oordeel te vellen over de vakbekwaamheid van beroepsbeoefenaren, doet zowel de vakbekwaamheid als de kennistoetsing tekort: vakbekwaamheid is meer dan kennis, en goede kennistoetsen rechtvaardigen ruimere gebruiksmogelijkheden.
5. Om de daadwerkelijke effectiviteit van nascholingsprogramma's die zich richten op kennistoename te kunnen vaststellen, is het noodzakelijk het kennisniveau ook geruime tijd nadat de nascholing is gevolgd, te peilen.
6. Competentietoetsen zouden aan kwaliteit winnen, als de energie besteed aan het zoeken naar de meest geëigende antwoordvorm, zou worden gebruikt voor het stellen van de meest geëigende vragen.
7. De literaire introductie in pocketboekuitgaven van klassieke Engelse romans, kan beter pas worden gelezen na lezing van het boek zelf.

8. Generaliseren mag in wetenschappelijke zin nastrevenswaard zijn, in het gewone dagelijks bestaan moet het worden afgeraden.
9. Alle pedagogisch didactische maatregelen die in het basisonderwijs worden doorgevoerd, zouden moeten worden ondersteund door een onderwijsbeleid dat is gericht op structurele vermindering van de klassegrootte. Zolang het zover niet is, dienen duo-banen en combinatieklassen te worden gestimuleerd.
10. In het onderwijs is de kwaliteit van de docent minstens zo belangrijk als de kwaliteit van het onderwijsprogramma. Bij het samenstellen van landelijk overdraagbare programma's ten behoeve van de huisartsgeneeskundige nascholing, moet daarom minstens zoveel aandacht worden geschonken aan de scholing van de docenten als aan de kwaliteit van het schriftelijke programmamateriaal.
11. De enorme toename van tekstverwerkingsfaciliteiten door middel van personal computers in het afgelopen decennium, heeft niet geleid tot meer efficiëntie of tijdbesparing bij het samenstellen van proefschriften.

Maastricht, Rijksuniversiteit Limburg, 1 december 1994.

KENNISTOETSING BIJ HUISARTSEN



KENNISTOETSING BIJ HUISARTSEN

1978

CIP GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Pollemans, Marjan

Kennistoetsing bij huisartsen / Marjan Pollemans. -

Maastricht : Universitaire Pers Maastricht

Proefschrift Maastricht. - Met lit. opg. - Met
samenvatting in het Engels.

ISBN 90-5278-159-1

Trefw.: huisartsgeneeskunde ; studietoetsen.

Productie: Datawyse | Universitaire Pers Maastricht

Omslag: de Vormgeverij, Maastricht

De uitgave van dit proefschrift kwam mede tot stand door een bijdrage van de Landelijke Huisartsen Vereniging.

KENNISTOETSING BIJ HUISARTSEN

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Rijksuniversiteit Limburg te Maastricht,
op gezag van de Rector Magnificus, Prof.dr. H. Philipsen,
volgens het besluit van het College van Dekanen,
in het openbaar te verdedigen op
donderdag 1 december 1994 om 16.00 uur

door

MARJAN POLLEMANS



UNIVERSITAIRE PERS MAASTRICHT

PROMOTORES

Prof.dr.mr. R.P.Th.M. Grol

Prof.dr. E. Schadé (Universiteit van Amsterdam)

CO-PROMOTOR

Dr. L.H.C. Tan (Universiteit Utrecht)

BEOORDELINGSCOMMISSIE

Prof.dr. W.H.F.W. Wijnen (voorzitter)

Prof.dr. H.F.J.M. Crebolder

Prof.dr. G.J. Kootstra

Prof.dr. J.C.M. Metz (Katholieke Universiteit Nijmegen)

Prof.dr. Th.B. Voorn (Universiteit Utrecht)

'Research is like motherhood. There is no such thing as an ugly baby.'

Anne Lennarson Greer.

The state of the art versus the state of the science. The diffusion of new medical technologies into practice. Int J Techn Assess Health Care 1988; 4: 5-26.

INHOUDSOPGAVE

Voorwoord	13
1 Inleiding tot het onderzoek	15
Samenvatting	15
1.1. Inleiding: achtergronden onderzoek	15
1.1.1. Maatschappelijke verantwoordingsplicht over kwaliteit van zorg	15
1.1.2. Kwaliteit in het huisartsgeneeskundig opleidingscontinuüm: stand van zaken	17
1.2. Literatuuroverzicht toetsing kennis in huisartsgeneeskundige nascholing	19
1.2.1. Rol (kennis)toetsing in deskundigheidsbevordering	19
1.2.2. Omschrijving kennisbegrip	21
1.2.3. Samenhang kennisniveau met achtergrondkenmerken	24
1.2.4. Nascholingsprogramma's en kenniseffecten	25
1.2.5. Vormen van kennistoetsing	28
1.2.6. Conclusie	31
1.3. Probleemstelling en doelen onderzoek	32
1.3.1. Probleemstelling	32
1.3.2. Onderzoekdoelen	34
1.4. Onderzoeksvragen	35
1.4.1. Ontwikkeling algemene huisartsgeneeskundige kennistoets	35
1.4.2. Samenhang kennisniveau en achtergrondkenmerken huisartsen	37
1.4.3. Ontwikkeling kennistoets ter bepaling effectiviteit nascholing	38
1.4.4. Invloed vraagvorm op huisartsgeneeskundige kennistoetsing	39
 Deel I: Ontwikkeling algemene kennistoets voor huisartsen	 41
2 Validiteit algemene huisartsgeneeskundige kennistoets	45
Samenvatting	45
2.1. Inleiding	46
2.2. Onderzoekopzet validiteitsstudie	47
2.2.1. Bepaling inhoudsvaliditeit	47
2.2.2. Bepaling constructvaliditeit	48
2.3. Blauwdrukontwikkeling	48
2.3.1. Methode blauwdrukontwikkeling	48
2.3.2. Resultaat blauwdrukontwikkeling	50
2.4. Empirische inhoudsvalidering blauwdruk	54
2.4.1. Methode: multivariate generaliseerbaarheidsanalyse	54
2.4.2. Resultaat multivariate generaliseerbaarheidsanalyse	55
2.5. Constructvalidering kennistoets	57
2.5.1. Methode	58
2.5.2. Resultaat: scoreverloop en expertiseontwikkeling	58
2.6. Conclusies validiteitsstudie algemene huisartsgeneeskundige kennistoets	59

3	Betrouwbaarheid algemene huisartsgeneeskundige kennistoets	61
	Samenvatting	61
3.1.	Inleiding	62
3.1.1.	Interpretatieperspectieven toetsscores	64
3.1.2.	Standaardmeetfout	66
3.2.	Onderzoeksopzet betrouwbaarheidsstudie	67
3.2.1.	Bepaling betrouwbaarheid individuele toetsscores	68
3.2.2.	Bepaling betrouwbaarheid groepsscores	68
3.3.	Resultaten betrouwbaarheidsstudie	70
3.3.1.	Betrouwbaarheid individuele toetsscores	70
3.3.2.	Betrouwbaarheid groepsscores	72
3.4.	Conclusies betrouwbaarheidsstudie algemene huisartsgeneeskundige kennistoets	73
4	Praktische toepasbaarheid algemene huisartsgeneeskundige kennistoetsing	77
	Samenvatting	77
4.1.	Inleiding	77
4.2.	Methode bepaling praktische toepasbaarheid	78
4.3.	Resultaten	79
4.3.1.	Acceptabiliteit	79
4.3.2.	Praktische uitvoerbaarheid	81
4.4.	Conclusies praktische toepasbaarheid algemene huisartsgeneeskundige kennistoets	84
	Deel II: Onderzoek naar samenhang kennisniveau en achtergrondkenmerken huisartsen	87
5	Algemeen kennisniveau in relatie tot achtergrondkenmerken van ervaren huisartsen	88
	Samenvatting	88
5.1.	Inleiding	89
5.2.	Methode onderzoek samenhang kennis en achtergrondkenmerken	89
5.2.1.	Opzet onderzoek	89
5.2.2.	Onderzoeksinstrumenten en variabelen	90
5.2.3.	Onderzoeksgroep	91
5.2.4.	Analyses	94
5.3.	Resultaten samenhang algemene huisartsgeneeskundige kennis en achtergrondkenmerken	95
5.3.1.	Gemiddelde kennistoetsscores	95
5.3.2.	Samenhang achtergrondkenmerken en kennisniveau	98
5.4.	Conclusies kennisniveau en achtergrondkenmerken huisartsen	103

6	Zelfkennis ervaren huisartsen	105
	Samenvatting	105
6.1.	Inleiding	105
6.2.	Methode bepaling samenhang kennis en oordeel over eigen expertise	106
6.3.	Resultaat samenhang kennis en oordeel over eigen expertise	108
6.4.	Conclusie zelfkennis huisartsen	110
	Deel III: Onderwerpgebonden kennistoetsing	113
7	Validiteit onderwerpgebonden kennistoetsen	115
	Samenvatting	115
7.1.	Inleiding	116
7.2.	Inhoudsvalidering	116
7.2.1.	Kennistoets vaginaal bloedverlies	117
7.2.2.	Kennistoets cholesterol	117
7.3.	Constructvalidering	118
7.3.1.	Kennistoets vaginaal bloedverlies	119
7.3.2.	Kennistoets cholesterol	124
7.4.	Conclusies validiteit onderwerpgebonden kennistoetsen	131
8	Betrouwbaarheid onderwerpgebonden kennistoetsen	134
	Samenvatting	134
8.1.	Inleiding	134
8.2.	Onderzoeksopzet betrouwbaarheidsstudie onderwerpgebonden kennistoetsen	135
8.3.	Resultaten	136
8.3.1.	Betrouwbaarheid individuele scores onderwerpgebonden kennistoetsen	136
8.3.2.	Betrouwbaarheid gemiddelde groepsscores onderwerpgebonden kennistoetsen	137
8.4.	Conclusies betrouwbaarheid onderwerpgebonden kennistoetsen	138
9	Praktische toepasbaarheid onderwerpgebonden kennistoetsing	140
	Samenvatting	140
9.1.	Inleiding	140
9.2.	Methode bepaling praktische toepasbaarheid onderwerpgebonden kennistoetsen	141
9.3.	Resultaten	141
9.3.1.	Acceptabiliteit onderwerpgebonden kennistoetsing	141
9.3.2.	Uitvoerbaarheid onderwerpgebonden kennistoetsing	142
9.4.	Conclusies praktische toepasbaarheid onderwerpgebonden kennistoetsing	145

Deel IV: Vraagvorm en kwaliteit toetsen	147
10 Invloed van de vraagvorm op de huisartsgeneeskundige kennistoetsing	148
Samenvatting	148
10.1. Inleiding	148
10.2. Methode vergelijkend onderzoek naar verschillende vraagvormen	149
10.2.1. Onderzoeksopzet vergelijkingsstudie	149
10.2.2. Onderzoeksinstrumenten	152
10.2.3. Onderzoeksgroep	153
10.3. Resultaten	154
10.3.1. Validiteit verschillende vraagvormen	154
10.3.2. Betrouwbaarheid verschillende vraagvormen	154
10.3.3. Praktische toepasbaarheid verschillende vraagvormen	156
10.4. Conclusies invloed vraagvorm	157
11 Beschouwing en discussie	159
Samenvatting	159
11.1. Inleiding	159
11.2. Algemene huisartsgeneeskundige kennistoetsing	161
11.3. Onderwerpgebonden kennistoetsing	165
11.4. Aanbevelingen	167
11.4.1. Beleidsaanbevelingen	167
11.4.2. Aanbevelingen voor verder onderzoek	168
Samenvatting	171
Summary	178
Literatuur	185
Bijlagen	203
1. Vragenlijst betreffende de inhoud en structuur van de huisartsgeneeskundige kennistoets	
2. Definities nevenindelingen blauwdruk	
3. Generaliseerbaarheidstheorie	
4. Formules betrouwbaarheidsstudies	
5. Voorbeeldpagina algemene huisartsgeneeskundige kennistoets februari 1992	
6. Vragenlijst achtergrondkenmerken	
7. Satisfactievragenlijst landelijke kennistoets februari 1992	
8. Instructiebrief huisartsen-deelnemers landelijke kennistoets februari 1992	
9. Voorbeeld individuele feedback kennistoets februari 1992, plus toelichtende brief	
10. Voorbeeldpagina kennistoets vaginaal bloedverlies	

11. Inventarisatie kenniselementen NHG-Standaard Cholesterol
12. Voorbeeldpagina kennistoets cholesterol
13. Satisfactievragenlijst onderwerpgebonden toetsen
14. Instructiebrief huisartsen-deelnemers onderwerpgebonden toetsing
15. Voorbeeld individuele feedback onderwerpgebonden toetsing, plus toelichtende brief
16. Retentievragenlijst kennistoets over cholesterol
17. Voorbeeldpagina kennistoets cholesterol (meerkeuzevragen)
18. Voorbeeld individuele feedback meerkeuzetoets, plus toelichtende brief

Curriculum vitae

VOORWOORD

Dit proefschrift gaat over toetsing van de kennis van huisartsen. Het project 'Kennistoetsing bij huisartsen' vormde de basis voor het onderzoek waarover wordt gerapporteerd. Vanwege de toenemende vraag naar verantwoorde en praktisch goed bruikbare instrumenten, die de kwaliteits- en deskundigheidsbevordering van huisartsen kunnen ondersteunen, is dit project, op aanvraag van het Nederlands Huisartsen Genootschap (NHG) en de Landelijke Huisartsen Vereniging (LHV), van november 1990 tot en met november 1993 door het Ministerie van Welzijn, Volksgezondheid en Cultuur gesubsidieerd. De Werkgroep Onderzoek Kwaliteit Huisartsgeneeskunde (WOK) te Maastricht was, tezamen met het Samenwerkingsverband van de Universitaire Huisartsopleidingen (SVUH) te Utrecht, verantwoordelijk voor de uitvoering van het onderzoek.

Het onderzoek is in nauwe samenwerking verricht met het onderzoek dat door Yvonne van Leeuwen aan de Rijksuniversiteit Limburg is uitgevoerd naar factoren die de kennisontwikkeling beïnvloeden van huisartsen-in-opleiding. Ook over dit laatste onderzoek is een proefschrift in voorbereiding. Aan beide onderzoeken lag de methode voor de toetsing van kennis ten grondslag die is ontwikkeld door het SVUH.

Het proefschrift is een typisch samenwerkingsproduct: het is het resultaat van een samenwerkingsproject met een Samenwerkingsverband. In de voorgaande alinea's zijn vijf instellingen genoemd die op enigerlei wijze bij het onderzoek waren betrokken. Het aantal personen dat bij het onderzoek was betrokken, was een veelvoud van dit getal. Al deze betrokkenen ben ik zeer veel dank verschuldigd, met name de huisartsen die aan de toetsen deelnamen. Zonder hun medewerking had het onderzoek niet kunnen worden uitgevoerd en was het proefschrift nooit tot stand gekomen. De huisartsen die deelnamen blijven noodzakelijkerwijs anoniem, maar een aantal andere betrokkenen wil ik hier graag noemen. Mijn beide promotoren, Richard Grol en Bert Schadé, en mijn co-promotor, Lisa Tan, wil ik danken voor de altijd enthousiasmerende en stimulerende manier waarop ze hun commentaar op steeds nieuwe stapels concept-hoofdstukken verpakten. De bijeenkomsten met jullie leidden steeds tot meer helderheid in de teksten en onderzoeksgegevens, en hadden voor mijzelf als bijzondere meerwaarde dat ik elke keer weer met plezier aan een nieuw concept begon.

Hiervoor is al aangegeven dat het onderzoek in nauwe samenwerking met Yvonne van Leeuwen is uitgevoerd. Beiden werden we met onze onderzoeken inhoudelijk ondersteund door Saskia Mol, en methodologisch en onderzoekstechnisch door Herman Düsman.

Yvonne, je creatieve denkvermogen is onovertroffen, en je bezit de zeldzame gave je niet geringe praktische huisartsgeneeskundige kennis en vaardigheden, te kunnen combineren met het vermogen tot reflectie over de voorwaarden die de kwaliteit van het huisartsenvak bepalen. Ik ben er zeer jaloers op geweest en ik heb er geweldig van geprofiteerd.

Saskia, je was fantastisch. Zo'n ondersteuning wens ik elke promovenda/dus toe. Ik heb er ook niets van gemerkt dat je aanstelling al een jaar geleden was afgelopen.

Herman, je was steeds bereid hulp te bieden en opnieuw uitleg te verschaffen over de technische en methodologische aspecten van het onderzoek. En je bleef vriendelijk! Je had altijd gelijk, maar soms merkte ik het pas véél later.

In de beginfase liep het onderzoek parallel met het onderzoek van Yvonne van Leeuwen. Voor de tijd en de aandacht die Riet Drop en Cees van der Vleuten in die periode ook aan mijn onderzoeksdeel bereid waren te schenken, dank ik hen van harte. Cees, ook nog veel dank voor je bereidheid om in de periode daarna mee te blijven denken over de meest wenselijke opzet van de betrouwbaarheidsstudies en voor het commentaar op enkele concept-hoofdstukken.

De leden van de beoordelingscommissie dank ik voor hun bereidheid om het manuscript van het proefschrift gedurende de (hete) zomervakantie te lezen.

Naast de direct bij het onderzoek betrokkenen, ben ik ook veel dank verschuldigd aan iedereen die meewerkte aan het creëren van de randvoorwaarden waarbinnen het onderzoek kon plaatsvinden: alle huisartsen/staffleden van de verschillende huisartsopleidingen die toetsvragen maakten;

de Werkgroep Kennis van het SVUH die onder het coördinatorschap van Just Eekhof verantwoordelijk was voor de definitieve vorm en formulering van de toetsvragen;

Jeroen Pielage van het SVUH, voor het ontwerpen van de programmatuur waarmee de verschillende vragenlijsten konden worden verwerkt, en Hans Eeckhout van het SVUH, voor het nauwgezet invoeren van de data en voor het verzorgen van de feedback naar de deelnemers aan de toetsen;

Karel Blanken van de afdeling deskundigheidsbevordering van de LHV, die zeer actief meewerkte aan de werving van huisartsen voor deelname aan de toetsing;

Marianne Hollander van het NHG en Trudy van der Weijden van de vakgroep Huisartsgeneeskunde van de Rijksuniversiteit Limburg, met wie de onderwerpgebonden toetsingen gezamenlijk zijn uitgevoerd;

de huisartsengroepen, die bereid waren deel te nemen aan de nascholingsprogramma's; het Dagelijks Bestuur van het Samenwerkingsverband, dat mij de laatste twee jaar de ruimte bood om mij geheel aan het onderzoek te wijden.

Tenslotte wil ik nog graag Elbrig Pasma en Paula Jobse van het SVUH noemen: altijd bereid nog even een klusje te klaren, een vragenlijst mooi vorm te geven, bestanden te achterhalen, brieven (met bijlagen!) te versturen, en blijmoedig alle andere administratieve werkzaamheden op zich te nemen die het leven van een proefschriftschrijfster vergemakkelijken.

De nauwe betrokkenheid van zoveel verschillende mensen heeft geleid tot een proefschrift, dat veel beter is dan ik in mijn eentje ooit tot stand had kunnen brengen.

Marjan Pollemans
Tilburg, 1994

Hoofdstuk 1

INLEIDING TOT HET ONDERZOEK

SAMENVATTING

Dit onderzoek betreft de waarde en toepasbaarheid van kennistoetsen binnen de deskundigheidsbevordering van huisartsen, en de eisen waaraan kennistoetsen daarbij moeten voldoen. Aan toetsingsgegevens, dat wil zeggen objectieve informatie over de capaciteiten en handelingen van huisartsen, wordt in het kwaliteitsbeleid van de beroepsgroep van huisartsen, een belangrijke ondersteunende rol toegekend (LHV, 1987, 1990, 1993-I, 1993-II; NHG, 1987, 1993-I). Dit geldt zowel op individueel niveau als op het niveau van de hele beroepsgroep (NHG, 1993-II). Op het niveau van de individuele huisarts wordt verwacht dat gegevens uit objectieve toetsing, goede hulp kunnen bieden bij de inrichting van het individuele nascholingstraject en bij het continue leerproces dat huisartsen gedurende de periode dat zij een praktijk uitoefenen, moeten doormaken. Op collectief niveau wordt verwacht dat toetsingsgegevens informatie opleveren die richting kan geven aan de opzet en planning van de nascholing. Toetsing moet er bovendien toe leiden dat de samenleving voldoende garantie wordt geboden voor een voortdurende vakbekwaamheid.

Het onderzoek beperkt zich tot één competentie-aspect, namelijk huisartsgeneeskundige kennis. Het beschikken over een adequaat kennisniveau wordt beschouwd als een noodzakelijke voorwaarde (uiteraard niet de enige voorwaarde) voor het kunnen leveren van zorg van goede kwaliteit. Van huisartsen mag daarom worden verwacht dat ze hun competentie op peil houden.

In dit hoofdstuk wordt eerst het kader geschetst waarbinnen het onderzoek is uitgevoerd. Vervolgens wordt de literatuur over de toetsing van kennis bij huisartsen samengevat. Daarna worden de doelen van het onderzoek geformuleerd, en de vraagstellingen toegelicht. De ordening van de vraagstellingen komt overeen met de ordening die in de literatuur is aangebracht. Deze ordening vormt de leidraad voor de opzet van het gehele proefschrift.

1.1. INLEIDING: ACHTERGRONDEN ONDERZOEK

1.1.1. Maatschappelijke verantwoordingsplicht over kwaliteit van zorg

Verantwoording over de *kwaliteit van de geleverde zorg*, wordt nationaal en internationaal in toenemende mate beschouwd als een maatschappelijke verplichting van al degenen die werkzaam zijn in de gezondheidszorg (KNMG Commissie Kwaliteit, 1990; Grol et al, 1993). Hieraan liggen verschillende motieven ten grondslag: de wens tot rationele sturing van de geleverde zorg, de relatieve schaarste van middelen, de toegenomen mondigheid van

patiënten, de introductie van concurrentie-principes in de gezondheidszorg en de verdere professionalisering van de beroepsgroep (Harteloh & Casparie, 1991). Na een eerste fase van expansie en een tweede van kostenbeheersing, wordt in de moderne, na-oorlogse gezondheidszorg het afleggen van verantwoording aan de samenleving over de geleverde zorg, een derde fase genoemd (Relman, 1988). De bereidheid rekenschap af te leggen wordt beschouwd als een maatstaf voor professionele verantwoordelijkheid (Benson, 1991). Kritische evaluatie van het (medisch) handelen moet in deze visie uiteindelijk leiden tot verbetering van de zorg (Gellhorn, 1991).

Aan het begrip kwaliteit van zorg wordt geen absolute betekenis gehecht. Kwaliteit van zorg wordt daarentegen vergelijkenderwijs vastgesteld. Zorg heeft meer kwaliteit naarmate de feitelijk geboden zorg beter voldoet aan de verwachtingen die de verschillende betrokkenen ten aanzien van die zorg hebben (Grol, 1991, 1993). Die verwachtingen kunnen per belanghebbende groepering verschillend zijn. Het begrip kwaliteit wordt dan ook omschreven als het vermogen om bepaalde doelen te bereiken. In haar advies inzake de 'Kwaliteit van Zorg' definieert de Raad voor Gezondheidsonderzoek het begrip kwaliteit als 'de mate waarin het geheel van eigenschappen van een produkt, proces, of dienst voldoet aan de eraan gestelde eisen, welke voortvloeien uit het gebruiksdoel' (RGO, 1990). Daarmee vormen de doeltreffendheid en doelmatigheid waarmee zorg wordt geleverd, belangrijke kenmerken (Donabedian, 1980).

Kwaliteitsbevordering omvat alle activiteiten die worden ondernomen om een bepaald kwaliteitsniveau van de zorg te handhaven of te bereiken (RGO, 1990). Een systematische aanpak hiervan wordt zowel structureel als op individueel niveau nagestreefd. Een wezenlijk element is regelmatige kwaliteitsbeoordeling op basis van toetsingsgegevens. Doel hiervan is gerichte actie te kunnen ondernemen ter verbetering van de kwaliteit van de zorg. Kwaliteitsbevordering wordt daarom ook wel onderscheiden in een diagnostisch deel: kwaliteitsmeting, en een therapeutisch deel: kwaliteitsverbetering (Casparie, 1989). De achterliggende gedachte is dat een eenmaal verleende kwalificatie die levenslang meegaat niet in het belang is van de samenleving, vanwege de veranderingen in medische kennis en in de sociale structuur van de samenleving (Lowenthal, 1981; Mueller, 1985). De aandacht voor systematische kwaliteitsbeoordeling op basis van toetsingsgegevens als een structureel aspect van de loopbaanontwikkeling, wordt volgens andere auteurs ook ingegeven door de wens tot handhaving van standaarden van zorg, toenemende publieke belangstelling en bewustzijn, en door de mogelijkheid van overheidsregelgeving (Norcini & Dawson-Saunders, 1994).

Kwaliteitsbeoordeling op basis van toetsingsgegevens wordt *intern* genoemd als organisaties van beroepsbeoefenaren of instellingen deze activiteit onder eigen verantwoordelijkheid in eigen beheer uitvoeren, en *extern* als een onafhankelijke instantie zoals de overheid, certificatie-instituten of verzekeraars, deze activiteit verrichten (RGO, 1990).

In Nederland hebben instellingen en beroepsbeoefenaren afspraken gemaakt met verzekeraars, patiëntenorganisaties en overheid over het te voeren kwaliteitsbeleid. Een van deze afspraken behelst het opzetten van *kwaliteitssystemen* (Sluijs & de Bakker, 1992). Deze moeten ertoe leiden dat de kwaliteit van zorg systematisch bewaakt, beoordeeld en verbeterd kan worden. De beroepsbeoefenaren hebben daarmee zelf de verantwoording op zich genomen voor de bewaking van de kwaliteit van zorg. Het betreft hier dus een vorm van interne kwaliteitsbeoordeling.

Formele registratie en herregistratie is een vorm van externe kwaliteitsbeoordeling, waarbij de samenleving via een instantie die daartoe is geautoriseerd, aan individuen toestaat, resp. blijft toestaan, een bepaald beroep uit te oefenen en een bepaalde titel te voeren (Boeringa & Sluijs, 1993). Deze toestemming wordt verleend nadat men ervan is verzekerd dat de kandidaat voldoet aan tevoren vastgestelde kwalifikaties en de minimale competentie bezit die noodzakelijk is om de openbare gezondheid, veiligheid en welzijn te waarborgen. Volgens Benson (1991) zou een herregistratieproces vier doelen moeten nastreven om te kunnen voldoen aan de behoefte tot het afleggen van verantwoording: verbeteren van de patiëntenzorg, standaarden vaststellen voor de praktijkvoering, levenslang leren bevorderen en het publiek, inclusief de wetgever, geruststellen door de competente arts te identificeren.

1.1.2. Kwaliteit in het huisartsgeneeskundig opleidingscontinuüm: stand van zaken

In Nederland heeft de aandacht voor de kwaliteit van de zorg geleid tot een aantal belangrijke ontwikkelingen op het terrein van de opleiding en de nascholing van huisartsen. De deskundigheidsbevordering van huisartsen wordt daarbij beschouwd als een continu doorlopend proces waarin, idealiter, het onderwijs wordt gestuurd door gegevens uit gerichte toetsing. In beide 'onderdelen' van het huisartsgeneeskundig opleidingscontinuüm, de opleiding tot huisarts en de nascholing of deskundigheidsbevordering, is dit cyclische proces van toetsing en scholing waarneembaar.

Het LHV-Basistakenpakket van de huisarts (Springer, 1983) vormt de inhoudelijke richtlijn voor de *huisartsopleiding* (Dubois et al, 1987). De huisartsopleiding is gebaseerd op een vaste structuur en op doelstellingen die door alle acht opleidingsinstituten zijn geaccepteerd (Grol et al, 1987). Ter bevordering en bewaking van de kwaliteit van de huisartsopleiding is onder meer gekozen voor de ontwikkeling van een landelijk evaluatiesysteem van de opleiding, en voor landelijke toetsing van de kennis en vaardigheden van aanstaande huisartsen (Pollemans e.a., 1988; Pollemans & Tan, 1988, 1990). Om dit doel te bereiken hebben de afdelingen huisartsopleiding van de acht vakgroepen huisartsgeneeskunde zich verenigd in een Samenwerkingsverband. Het instrument voor de toetsing van de kennis van aanstaande huisartsen, is het meest ver ontwikkelde, landelijk toepasbare instrument. Dit instrument heeft de vorm van een voortgangstoets, waarmee de kennisontwikkeling van de aanstaande huisartsen periodiek in kaart kan worden gebracht (Pollemans e.a., 1991).

Ook op het gebied van de *nascholing* van huisartsen vinden ontwikkelingen plaats die tot doel hebben de kwaliteit van de door de huisarts geleverde zorg te verbeteren (LHV, 1986, 1987, 1990, 1993-I, 1993-II). De beroepsverenigingen gaan er daarbij vanuit dat een sterkere positie van de huisarts in de eerste lijn ten dienste staat van de kwaliteit van de gezondheidszorg als geheel, en zal leiden tot een evenwichtigere ontwikkeling van de gezondheidszorg (LHV, 1986). Dat impliceert wel dat garanties moeten kunnen worden gegeven voor de kwaliteit van de geleverde huisartsgeneeskundige zorg (LHV, 1987). Door de Landelijke Huisartsen Vereniging (LHV) en het Nederlands Huisartsen Genootschap (NHG) is daarom een gezamenlijk *kwaliteits- en deskundigheidsbevorderingsbeleid* opgesteld, dat moet leiden tot daadwerkelijke veranderingen in het handelen en in de praktijkvoering door huisartsen (SDH, 1990). Dit moet uiteindelijk leiden tot een huisartsgeneeskundig *kwaliteitssysteem* waarin patiënten, beleidsmakers en huisartsen zich kunnen herkennen. Dit systeem kent drie pijlers: richtlijnen voor de praktijk voor verantwoord huisartsgeneeskundig handelen, zoals worden

ontwikkeld in de vorm van de NHG-Standaarden, kwaliteitstoetsing, en kwaliteitsverbetering door deskundigheidsbevordering (Grol, 1991). Het kwaliteitsbeleid is begin jaren negentig vastgesteld en wordt uitgewerkt in een aantal vijfjaarsperioden, waarin steeds een aantal verdere stappen wordt gezet (Rutten & Thomas, 1993).

Het kwaliteitsbeleid bevat een aantal kernpunten. Deze betreffen onder meer de verlenging van de huisartsopleiding tot drie jaar, een duidelijker structurering van de nascholing die tot minder vrijblijvendheid moet leiden, het ontwikkelen van Standaarden met daarop gebaseerde deskundigheidsbevorderingspakketten, en de ontwikkeling van toetsingsinstrumenten voor het bepalen van de effectiviteit van nascholing. Daarnaast past verplichte herregistratie, op basis van geaccrediteerde inhoudelijke criteria, in het kwaliteitsbeleid.

De *NHG-Standaarden* voor het huisartsgeneeskundig handelen, spelen bij de uitwerking van het kwaliteitsbeleid een essentiële rol (NHG, 1987, 1989; Rutten & Thomas, 1993). Standaarden vormen bundelingen van criteria en richtlijnen die zodanig zijn gestructureerd en gerangschikt dat ze met elkaar, bij de huidige stand van de wetenschap, de best mogelijke medische eerstelijnszorg beschrijven (NHG, 1987). De formulering ervan heeft betrekking op het te volgen diagnostisch en therapeutisch beleid bij een bepaalde klacht, ziekte of aandoening, maar betreft ook eisen die worden gesteld aan bij voorbeeld medische registratie, administratie, instrumentarium of praktijktoegankelijkheid (NHG, 1987, 1989). NHG-Standaarden worden gezien als richtlijnen, waarvan in specifieke individuele gevallen beargumenteerd kan worden afgeweken (Rutten & Thomas, 1993). Met de Standaarden wordt de kwaliteit van handelen van de beroepsgroep van huisartsen naar buiten toe verantwoord. Ze zijn een uiting van de professionalisering van de huisartsgeneeskunde en een middel om onaanvaardbare interdoktervariatie terug te dringen. Omdat ze een nadere concretisering van het Basistakenpakket bieden, moeten ze ten slotte een basis vormen voor toetsing en nascholing (Rutten & Thomas, 1993). Om dit laatste doel te bereiken, ontwikkelt het NHG deskundigheidsbevorderingspakketten, die zijn gebaseerd op de Standaarden en zijn bestemd voor gebruik binnen de nascholing. Om het leereffect van deze pakketten te vergroten, worden ze voorzien van korte (kennis)toetsen (NHG, 1993-I; LHV, 1993-II).

Tenslotte bestaan er plannen om op termijn, in het kader van de *herregistratie*, verplichte deelname van huisartsen aan formele toetsing te introduceren (LHV, 1990). Daarbij wordt als uiteindelijk ideaal gezien een situatie waarbij huisartsen alleen worden geherregistreerd, als zij voldoen aan een aantal kwalitatieve eisen. Huisartsen zouden, onder meer op basis van adequate toetsingsgegevens, moeten kunnen aantonen dat zij aan deze eisen voldoen (van Leeuwen e.a., 1991). Huisartsen zouden, bij gebleken tekortkomingen, verplicht zijn deel te nemen aan geaccrediteerde deskundigheidsbevordering, specifiek gericht op hun lacunes.

Het kwaliteitsbeleid van de beroepsgroep richt zich dus op een systematische kwaliteitsbevordering, gebaseerd op een cyclisch proces van toetsing en scholing. Daarbij wordt geformaliseerde en verplichte deskundigheidsbevordering voorgestaan, die onder meer is gebaseerd op objectief verkregen toetsings- en evaluatiegegevens.

1.2. LITERATUUROVERZICHT TOETSING KENNIS IN HUISARTSGENEESKUNDIGE NASCHOLING

De belangrijke plaats die in het kwaliteitsbeleid aan toetsingsgegevens wordt toegekend, maakt de vraag van belang wat er in de literatuur is te vinden over de toepassing van (kennis)toetsingsgegevens op dit gebied. De literatuur wordt samengevat vanuit verschillende invalshoeken. Eerst staan de gebruiksfuncties van toetsingsgegevens voor nascholingsdoelen centraal: *waarom* wordt er getoetst? Vervolgens wordt ingegaan op het kennisbegrip: *wat* wordt er getoetst? Na deze vooral theoretisch georiënteerde paragrafen volgen enkele paragrafen met resultaten van empirisch onderzoek. Samengevat wordt welke ervaringen zijn beschreven met daadwerkelijke toepassing van kennistoetsen. Daarbij wordt ingegaan op wat er bekend is over de samenhang tussen achtergrondkenmerken van huisartsen en hun kennisniveau. In de daarop volgende paragraaf worden de bevindingen weergegeven bij toepassing van kennistoetsen in nascholingseffectstudies. Tenslotte wordt de literatuur beschreven vanuit het perspectief van de toetsvorm: het *hoe* van kennistoetsing.

1.2.1. Rol (kennis)toetsing in deskundigheidsbevordering

De eerste invalshoek van waaruit de literatuur wordt beschreven, betreft de gebruiksfuncties die aan toetsingsgegevens in de nascholing worden toegekend.

Nascholingsprogramma's voor huisartsen beogen (nieuwe) inzichten op relevante gebieden zodanig over te dragen dat deze uiteindelijk door huisartsen in hun praktijkvoering tot uiting worden gebracht (Davis et al, 1990). Het uiteindelijke doel is verbetering van de zorg aan de patiënt (Miller, 1987; McLaughlin & Donaldson, 1991; Lawrence & Pritchard, 1992). Het praktijkgedrag (*performance*: datgene wat de arts daadwerkelijk in de dagelijkse praktijk doet) is een cruciale factor om te kunnen beoordelen of de nascholing doeltreffend is geweest. De kennis waarover een huisarts beschikt, is niet direct waarneembaar, maar vormt een aspect van diens *competentie* of vermogen tot handelen: datgene waartoe een arts in staat is. Adequate competentie is geen garantie voor een adequate performance, maar wordt wel beschouwd als een noodzakelijke voorwaarde (Kane, 1986). Competentie wordt vastgesteld in test- of toetssituaties die zijn gericht op het bepalen van kennis en vaardigheden. In de praktijk bepalen ook de efficiëntie en de consultduur de performance (Rethans, 1991). Aangenomen mag echter worden dat artsen die niet beschikken over de vereiste kennis en vaardigheden deze ook in de praktijk niet kunnen toepassen.

Deskundigheidsbevordering is onderdeel van het medische onderwijscontinuüm. Vrijwel alle gevestigde huisartsen erkennen dat ze een professionele verantwoordelijkheid hebben tot continu leren (Forrest et al, 1989). De nascholing van artsen onderscheidt zich in een aantal opzichten van hun beroepsopleiding (Suter et al, 1981). Er treedt een verschuiving op van een institutionele naar een individuele verantwoordelijkheid voor de invulling van het onderwijs. Het leren is niet meer gericht op het bereiken van een specifieke kwalificatie, maar op het wegnemen van in de praktijk ervaren deficiënties. Het georganiseerde en continue leren wordt meer geïndividualiseerd en episodischer van karakter. De eigen motivatie tot leren speelt een belangrijke rol (Lawrence & Pritchard, 1992). Van ervaren (huis)artsen wordt verwacht dat ze in staat zijn hun eigen leerbehoeften te definiëren, hun eigen leerdoelen te stellen, leerbenaderingen te kiezen, evaluatiemethoden te selecteren en uiteindelijk de verworven competentie toe te passen (Suter et al, 1984). Daarbij past een toetsings- of evaluatiebeleid

dat is gericht op continue verbetering (Berwick, 1989). Volgens deze benadering moeten artsen gedurende hun gehele beroepsmatige bestaan nieuwe kennis en vaardigheden verzamelen, ongeacht hun competentieniveau (Hays et al, 1993). Deze kennis en vaardigheden moeten worden verworven door gerichte nascholing. Systematisch verzamelde toetsingsgegevens kunnen daarbij een belangrijke rol spelen. Op verschillende abstractieniveaus kunnen deze bijdragen aan een meer *effectieve en efficiënte deskundigheidsbevordering* (Anonymus, 1987). Dit betreft zowel het bevorderen van een doelgerichte planning van de nascholing, als het bijdragen aan een objectiever inzicht in de sterke en zwakke kanten van de individuele huisarts (Grol & Zwaard, 1990). Dit impliceert dat aan toetsingsgegevens een educatieve of diagnostische functie wordt toegekend (Ewell, 1991), zowel op groepsniveau (screening ten behoeve van adequate planning) als op het niveau van de individuele huisarts. De educatieve functie houdt een verbreding in van de visie op toetsing. Toetsing wordt niet alleen beschouwd als een vorm van externe beoordelingsprocedure, maar wordt vanuit deze visie vooral beschouwd als een onderdeel van het curriculum, het onderwijs en het leerproces (Brown, 1992). Het betekent onder meer dat veel aandacht moet worden gegeven aan de verantwoording van de inhoud van de toetsen en de toetsprocedures, en aan de betekenis van de toetsresultaten.

Een onomstreden uitgangspunt voor deskundigheidsbevordering van artsen is dat deze moet aansluiten bij persoonlijk ervaren tekortkomingen, c.q. zelf onderkende leerbehoeften (Stein, 1981; Means, 1984; Millac, 1985; Miller, 1987; Parry, 1990; Stanley et al, 1993). Subjectieve leerbehoeften kunnen echter door een speciale belangstelling voor bepaalde onderwerpen worden ingegeven. Dat kan ertoe leiden dat (huis)artsen vooral voor nascholingsprogramma's kiezen over onderwerpen waar ze al redelijk veel van weten (Sibley et al, 1982). Ook zouden artsen hun eigen 'blinde vlekken' niet herkennen, waardoor ze geen behoefte aan nascholing op die terreinen ontwikkelen (Corbett, 1979; Manning et al, 1984; Manning & DeBakey, 1992). Als artsen vrij zijn om hun nascholingsprogramma zelf vast te stellen, bestaat dus het risico dat niet wordt gekozen voor onderwerpen waarvoor nascholing het meest nodig is. De zelf ervaren leerbehoeften worden aangeduid met de term '*perceived needs*'. Daartegenover staan '*true needs*' die meer objectief worden vastgesteld door onafhankelijke toetsing op basis van feitelijke gegevens (Commissie Nascholing NHG, 1989). Veel nascholingsprogramma's ontberen effectiviteit omdat ze niet aansluiten op deze '*true needs*' (Asbaugh & McKean, 1976; Laxdal, 1982; Davis, 1986). De relativiteit van het eigen oordeel komt ook naar voren in een onderzoek naar de behoeften en wensen van Nederlandse huisartsen ten aanzien van nascholingsprogramma's. Specifieke deskundigen en gewone huisartsen bleken onderling niet hetzelfde belang te hechten aan medisch-inhoudelijke nascholingsthema's. Tevens bleek dat huisartsen meer overeenstemden in hun mening over het belang van bepaalde thema's voor de gezondheidszorg in het algemeen, dan over het belang van deze thema's voor hun eigen kennis en vaardigheden (Bloemer & Poiesz, 1987).

Objectief bepaalde tekortkomingen hoeven dus niet samen te vallen met subjectieve leerbehoeften. De nascholing moet daarom zo zijn vormgegeven dat ze artsen ondersteunt in het achterhalen van hun werkelijke leerbehoeften (Levine et al, 1984-I). Dat betekent dat het aanbod aan nascholingsprogramma's tegemoet moet komen aan objectief vastgestelde lacunes. Tegelijkertijd betekent het overigens ook dat artsen objectief bepaalde tekortkomingen moeten onderkennen als subjectieve leerbehoeften. Hiervoor moet het belang van de tekortkomingen voor het praktijkgedrag duidelijk zijn (Grol & Zwaard, 1990; Mann & Chaytor, 1992).

1.2.2. Omschrijving kennisbegrip

De tweede invalshoek voor de literatuurbeschrijving wordt gevormd door het kennisbegrip dat in de toetsen naar voren komt. Toepassing van kennistoetsen ten behoeve van de deskundigheidsbevordering van huisartsen vereist explicitering van dit kennisbegrip. De toetsen moeten recht doen aan het abstractieniveau dat de kennis van ervaren huisartsen kenmerkt (Chi & Glaser, 1980). Daarvoor is inzicht nodig in de werking van het geheugen en in de wijze van probleemoplossen bij toenemende expertise (Glaser, 1990).

Binnen de cognitieve psychologie is de belangstelling voor expertise en de rol die kennis en ervaring daarin spelen, relatief recent. Deze belangstelling hangt samen met de opkomst van computersimulaties en onderzoek naar artificiële intelligentie (Glaser, 1984; Waldrop, 1984). Onderzoeksbevindingen maakten duidelijk dat als op een bepaald domein over relatief veel kennis wordt beschikt, dit samenhangt met een beter begrip en beter onthouden van nieuwe kennis binnen dat domein (Johnson, 1988). De expert blijkt zich vooral van de beginnende beroepsbeoefenaar te onderscheiden door een effectievere en efficiëntere structurering van zijn kennisbestand.

Tot voor enkele decennia werd probleemoplossen beschouwd als gekenmerkt door algemene redeneervaardigheden die op uiteenlopende taken kunnen worden toegepast (Sackett et al, 1985). De nieuwe bevindingen laten daarentegen zien dat kennis en probleemoplossen geen onafhankelijke cognitieve constructen zijn (Glaser, 1990). De betere structurering van het kennisbestand van experts ontstaat volgens deze inzichten door het ontwikkelen van specifieke probleemoplossingsstrategieën ten gevolge van toenemende ervaring met het oplossen van domeinspecifieke problemen (Johnson, 1988; Prawat, 1989). Probleemoplossen moet volgens deze theorie worden gezien als het toepassen van kennis binnen een bepaald domein. Het ontwikkelen van expertise wordt vervolgens beschouwd als identiek aan het leren oplossen van problemen binnen bepaalde domeinen (Boshuizen, 1989; Boshuizen & Schmidt, 1993). In een review van de literatuur over de interactie tussen kennis van het domein en van probleemoplossingsstrategieën (zogenaamde strategische kennis), concluderen Alexander & Judy (1988) dat een zekere mate van basale kennis nodig is om expertise te kunnen ontwikkelen. Inaccurate of incomplete kennis van het domein kan het leerproces hinderen of belemmeren. Nieuwe informatie kan daardoor bijvoorbeeld niet worden opgenomen en verwerkt. Personen met weinig kennis van een domein kunnen de juiste probleemoplossingsstrategieën niet toepassen, en neigen ertoe het gebied dat ze al beheersen uit te breiden, in plaats van te verkennen wat ze nog niet weten (Alexander & Judy, 1988). De conclusie is dat de kennisstructuur kwalitatief voldoende basis moet hebben om faciliterend te kunnen werken op de expertise-ontwikkeling (Messick, 1984-I).

Ook Snoek (1989), die onderzoek deed naar de strategieën die ervaren neurologen hanteren in hun denkproces, concludeert dat de ontwikkeling van inhoudsspecifieke probleemoplossingsstrategieën de cruciale factor vormt in het ontstaan van expertise. Volgens hem beschikken meer ervaren klinici niet alleen over meer kennis, maar 'redeneren' ze ook beter dan onervarenen. Redeneerprocessen worden door hem opgevat als het intuïtief herkennen, doorzien en hanteren van regels en hun uitzonderingen binnen een bepaald vakgebied. Het kennisbestand wordt door deze redeneerwijzen gestructureerd; beide nemen met uitbreiding van ervaring toe.

De expertisetheorie sluit aan op de theorie van Ebel (1982) over het kennisbegrip waarop toetsen die zijn bedoeld voor gebruik in opleidingssituaties, zich moeten richten. Volgens Ebel is informatie die alleen als zodanig wordt opgeslagen in het geheugen en die niet wordt verbonden met andere, al bekende zaken, de laagste en minst nuttige vorm van kennis. Door het denkproces wordt de informatie tot begrip en raakt geïntegreerd in een systeem van relaties tussen concepten en ideeën dat de structuur van kennis vormt. De cruciale vraag is vervolgens of het denkvermogen een algemene vaardigheid is, onafhankelijk van het probleem dat aan de orde is, of dat het een vermogen is dat specifiek is voor een bepaald soort van problemen. Ook volgens Ebel lijkt het erop dat het om dit laatste gaat: probleemoplossende vermogens neigen ertoe specifiek te zijn (Ebel, 1982).

De nieuwe theorie spoort tenslotte met visies op het ontstaan van expertise zoals die zijn ontwikkeld vanuit de praktijk van met name het onderwijs dat opleidt tot een bepaalde discipline (Everwijn e.a., 1990; Simons, 1990; Simons & Verschaffel, 1992). De combinatie en de integratie van domeinspecifieke vakkennis en probleemoplossende vaardigheden leidt tot professionele competentie. Daarbij worden eerdere leerervaringen zodanig in het geheugen geïncorporeerd, dat ze onder nieuwe omstandigheden kunnen worden gebruikt (Gick & Holyoak, 1987). Deze leerervaringen moeten zijn gebaseerd op specifieke en gevarieerde casuïstiek om 'transfer', het kunnen toepassen van eerder verworven kennis in nieuwe situaties, te bewerkstelligen (Spiro et al, 1987). In dit leerproces worden de algemene heuristieken die ten grondslag liggen aan bijvoorbeeld probleemoplossende vaardigheden, 'gecontextualiseerd', dat wil zeggen in verband gebracht met de vakspecifieke inhoud (Everwijn e.a., 1990). Volgens deze benadering is kennis en beheersing van het metier karakteristiek voor vakmanschap: meesterschap is vooral kennerschap (Everwijn & Palm, 1993).

De expertisetheorie kan als volgt worden samengevat: bij toenemende ervaring binnen een bepaald domein wordt het kennisbegrip coherenter en beter gestructureerd; kennis wordt nuttiger door een efficiënter gebruik, omdat toenemende ervaring de condities beter bekend maakt waaronder de kennis toepasbaar is. Experts ontwikkelen daardoor steeds doeltreffender probleemoplossings-strategieën om hun kennis op het juiste moment goed te kunnen gebruiken (Chi & Glaser, 1988; Glaser, 1990).

De nieuwe expertisetheorie heeft de theorievorming over de *ontwikkeling van medische expertise* sterk beïnvloed (Boshuizen, 1989; Schmidt & Boshuizen, 1993; Hobus, 1994). Medische expertise ontwikkelt zich volgens deze zienswijze via een aantal stadia (Bordage et al, 1990). In het eerste stadium is geen sprake van kennis of is deze ontoegankelijk vanwege de slechte organisatie van de kennis in het geheugen (*no knowledge* of *low knowledge*). De kennisontwikkeling begint met de opbouw van een conceptueel rijke en rationele kennisbasis die steunt op de basiswetenschappen. De kennis is nog gebrekkig in het geheugen opgeslagen (*dispersed knowledge*). In het volgende stadium kent de student bepaalde lijsten en kenmerken van prototypische ziektes. Deze kennis is netwerkwachtig gestructureerd opgeslagen waardoor globale representatie van problemen mogelijk is (*elaborated knowledge*). Tot en met dit stadium vindt het diagnostisch redeneren plaats door het afleiden van hypothesen uit de beschikbare gegevens en door vervolgens deze via een proces van logisch denken te verifiëren of te falsificeren. In het laatste stadium (*compiled knowledge*, later ook wel *kennisencapsulatie* genoemd) heeft de dan ervaren arts zich, via de opgedane eigen ervaring met individuele casuïstiek, een heel repertoire van ziektegevallen

eigen gemaakt (Feltovitch & Barrows, 1984; Boshuizen & Schmidt, 1993). Deze zijn in het geheugen georganiseerd als *illness scripts* of *instant scripts*. De uitgebreide kennisnetwerken in het geheugen uit de eerdere fasen, zijn nu ingekort op basis van eigen ervaring met patiëntcasuïstiek (Grant & Marsden, 1988). Diagnostiek gebeurt vooral via patroonherkenning (*pattern recognition*) en niet, zoals in de voorgaande stadia, alleen op basis van hypothetico-deductief redeneren (Patel & Groen, 1986; Boshuizen, 1989; Schmidt et al, 1990; van der Vleuten & Norman, 1991; Hobus, 1994).

Het laatste expertisestadium, dat van de gecompileerde of geëncapsuleerde kennis, waarin de ervaren arts diagnostiek bedrijft op basis van patroonherkenning, is kwalitatief verschillend van de voorgaande. Het is namelijk vooral gebaseerd op persoonlijke ervaringen met patiëntcasuïstiek bij de toepassing van de kennisconcepten die in de eerste drie stadia een rol spelen (Norman et al, 1989; Schmidt & Boshuizen, 1993). Medische expertise moet dan ook worden beschouwd als een verworvenheid die individuele verschillen vertoont, omdat het de persoonsgebonden integratie betreft van de individuele (praktijk)ervaring van artsen met de formeel verworven kennis (Grant & Marsden, 1987). De individuele ervaringen, vastgelegd in *illness scripts*, bevatten relatief weinig informatie over pathofysiologische oorzaken van klachten en ziektebeelden. Daarentegen bevatten ze wel veel praktijkrelevante informatie over ziekte, gevolgen, en de context waarbinnen de ziekte zich ontwikkelt (Boshuizen et al, 1987; Hobus e.a., 1988; Schmidt et al, 1990, 1992; Hobus, 1994). Ze worden door experts vooral in routine-situaties toegepast (Bordage & Zacks, 1984). Alleen bij complexe gevallen is het handelen (tevens) gebaseerd op een meer analytische oplossingsstrategie (Boshuizen et al, 1988; Schmidt & Boshuizen, 1993). Artsen op een intermediair niveau tussen beginners en experts in, met veel theoretische kennis maar zonder relevante praktijkervaring, gebruiken deze laatste oplossingsstrategieën, die zijn gebaseerd op pathofysiologie en causale redeneerprocessen, ook bij veel minder complexe situaties (Groen & Patel, 1985; Patel & Groen, 1986; Schmidt et al, 1990). Toepassing van deze strategie is echter minder efficiënt dan de patroonherkenningsstrategie (Boshuizen & Schmidt, 1993).

De kennis die ten grondslag ligt aan de verschillende probleemoplossingsstrategieën die artsen toepassen, worden door Schmidt et al (1990) samengevat in twee typen van kennis: *conceptuele* en *ervaringskennis*. Tussen beide vormen bestaat een kwalitatief verschil. Conceptuele kennis is de algemene, gemeenschappelijke kennis van het vakgebied, die wordt overgedragen in de opleiding. Het is de kennis die noodzakelijk is om op basis van een logisch causaal en analytisch redeneerproces hypothesen te kunnen afleiden. Ervaringskennis ofwel de 'gegeneraliseerde ervaring', gecombineerd met de herinnering aan concrete patiëntcasuïstiek (Schmidt et al, 1990, 1992), is de kennis die ten grondslag ligt aan de individuele *illness scripts* die het 'medisch geheugen' van de ervaren arts organiseren. Ervaringskennis wordt aangelegd op basis van conceptuele kennis (Grant & Marsden, 1987; Norman, 1988). Beide typen kennis bestaan in het geheugen van de individuele arts naast elkaar, waarbij met toenemende ervaring de hoeveelheid conceptuele kennis als het ware indikt (Schmidt et al, 1990, 1992).

Het theoretische model voor de ontwikkeling van medische expertise is in de praktijk getoetst in onderzoek naar verschillen in diagnostiekvorming tussen medische studenten in verschillende opleidingsstadia en ervaren artsen (Patel et al, 1986). De resultaten bevestigden de op de theorie gebaseerde verwachtingen. Experts identificeren de meest relevante

componenten uit klachten- of symptomencomplexen, maar geven geen gedetailleerde biomedische verklaringen. Deze laatste worden wel beschreven door studenten in de eindfase van hun medische opleiding. Zij herkennen echter niet alle relevante componenten uit klachten- of symptoombeschrijving (Groen & Patel, 1988; Arocha & Patel, 1990). Studenten lijken meer in termen van ziekten te denken, ervaren artsen in termen van patiënten (Claessen & Boshuizen, 1985). Andere bevindingen zijn dat de hoeveelheid conceptuele kennis niet lineair stijgt met toenemende ervaring, maar eerst, vanaf beginnersniveau, stijgt tot een zogenaamd intermediair niveau (waarop wel alle relevante conceptuele kennis wordt beheerst, maar nog weinig praktijkervaring bestaat) en vervolgens weer daalt met toenemende praktijkervaring (Grant & Marsden, 1988; Schmidt et al, 1988; Norman et al, 1989). Met betrekking tot de kwaliteit van de diagnostische hypothesevorming, werd gevonden dat deze samen lijkt te hangen met specifieke ervaring in de vorm van eerdere voorbeelden van patiëntencasus (Norman et al, 1989). Verder werd gevonden dat ervaren artsen betere diagnostische prestaties leveren dan minder ervaren artsen, omdat ze meer en intensiever gebruik maken van contextuele patiëntinformatie (Hobus, 1994).

Toetsen die algemeen toepasbaar moeten zijn in de deskundigheidsbevordering van huisartsen, moeten zich richten op de kennis die elke huisarts aan het eind van zijn opleiding wordt geacht te beheersen. Dit impliceert dat ervan wordt uitgegaan dat huisartsgeneeskundige kennis een ander conceptueel domein omvat, dan het kennisdomein dat aan het eind van de artsopleiding moet worden beheerst (Metz e.a., 1994).

Mede op basis van de hiervoor beschreven theorie over het ontstaan van medische expertise, wordt het kennisbegrip in algemeen toepasbare toetsen omschreven als *gegeneraliseerde huisartsgeneeskundige ervaringskennis*, die tevens de conceptuele kennis omvat die aan het eind van de huisartsopleiding moet worden beheerst. Het betreft de kennis waarop elke huisarts in beginsel aanspreekbaar is. Hoe deze kennis precies in het geheugen van elke individuele huisarts is opgeslagen, is minder relevant. De kennis kan geheel als conceptuele kennis zijn opgeslagen, maar ook in de vorm van illness scripts. Daarbij wordt ervan uitgegaan dat toenemende ervaring leidt tot een geconcentreerde en snel toegankelijke kennisbasis, waarin de meest elementaire inhoudelijke kenniselementen in samenhang met specifieke casuïstiek zijn opgeslagen. De vragen in kennistoetsen die gericht zijn op een dergelijk kennisbegrip, moeten een beroep doen op deze elementaire, huisartsgeneeskundig relevante kenniselementen. Deze worden ook wel aangeduid als zogenaamde '*key features*': specifieke en uiterst relevante elementen die een bepaald klinisch probleem of een casus kenmerken. Kennis over deze elementen is cruciaal om het probleem te kunnen oplossen. Key features kunnen betrekking hebben op de kritische (kennis)elementen die nodig zijn om een probleem te kunnen oplossen, maar ook op bekende misvattingen die tot een verkeerde probleemoplossing leiden of op in de praktijk moeilijk gebleken aspecten van diagnostiek en beleid (Bordage & Page, 1987).

1.2.3. Samenhang kennisniveau met achtergrondkenmerken

De paragrafen over de gebruiksfuncties van kennistoetsen en over het kennisbegrip dat in de toetsen naar voren zou moeten komen, waren voornamelijk theoretisch georiënteerd. In deze paragraaf wordt ingegaan op de informatie die de literatuur levert over empirisch onderzoek naar de factoren die met het kennisniveau van (huis)artsen samenhangen of het lijken te beïnvloeden.

Prestaties op toetsen worden zowel door persoons- als omgevingskenmerken beïnvloed (Messick, 1984-II). Hiermee moet bij de interpretatie van toetsresultaten rekening worden gehouden, omdat de validiteit van de bevindingen erdoor kan worden aangetast. Behalve individugebonden kenmerken, bestaan er wellicht ook structurele achtergrondkenmerken die samenhang vertonen met het kennisniveau van huisartsen. Inzicht in deze kenmerken kan richtinggevend zijn voor de nascholing voor bepaalde groepen huisartsen.

De literatuur laat zien dat het kennisniveau van artsen, gemeten met schriftelijke kennistoetsen, lijkt te dalen met hun leeftijd, of met de termijn die is verstreken sinds hun opleiding (Day et al, 1988; Ramsey et al, 1989; Steel et al, 1989; Salem-Schatz et al, 1990). Deze ervaring is vooral gebaseerd op de resultaten behaald op Amerikaanse recertificatie-examens bij internisten (Meskauskas & Webster, 1975; Norcini et al, 1985). Hetzelfde leeftijdseffect werd ook in recertificatie-toetsen gevonden die waren gericht op 'family practice' (Leigh et al, 1990; Spike & Veith, 1990). Volgens Norcini & Dawson-Saunders (1994) is een afname in prestaties op schriftelijke meetinstrumenten bij artsen te verwachten na een periode van 10 tot 15 jaar. De achteruitgang in kennis met toenemende leeftijd kan een gevolg zijn van het feit dat kandidaten minder kennis hebben, maar kan er ook mee te maken hebben dat examenvragen minder relevant zijn voor gevorderde practici. Geadviseerd wordt voor dit probleem een oplossing te zoeken door te streven naar een zo veel mogelijk op de praktijk afgestemde ('practice tailored') examen-inhoud (Norcini & Dawson-Saunders, 1994).

In een onderzoek van Day et al (1988) werd nagegaan of de zwakkere prestaties van oudere artsen (1947 internisten; vrijwillige participanten aan recertificatie-examen van 1980) konden worden verklaard door de mate van recente kennis in de gebruikte toetsen. Inderdaad bleek de groep die het meest recent was afgestudeerd het hoogst te scoren op nieuwe en veranderde kennis. Geconcludeerd werd dat de afnemende kennis bij oudere internisten mogelijk meer was gerelateerd aan de mogelijkheden van deze artsen om hun kennis up to date te houden en uit te breiden, dan met veranderingen in hun oorspronkelijk kennisbezit of in problemen met de schriftelijke multiple choice examenvorm. Tevens werd gevonden dat solisten gemiddeld lager scoorden dan artsen die in groepsverband samenwerkten (Day et al, 1988). Het kennisniveau van artsen blijkt niet alleen samen te hangen met leeftijd en praktijkkenmerken, maar ook met de tijd die wordt besteed aan onderwijs en onderzoek (Steel et al, 1989). Andere variabelen die worden geassocieerd met een hoger kennisniveau zijn een langere duur van de opleiding, deelname aan nascholingsactiviteiten en het lezen van vakbladen (Bertram & Brooks-Bertram, 1977).

1.2.4. Nascholingsprogramma's en kenniseffecten

In deze paragraaf wordt de literatuur samengevat over de effectiviteit van nascholing. Daarbij wordt speciaal ingegaan op wat er bekend is over de mate waarin aan toetsgegevens een indicatie voor effectiviteit mag worden gehecht.

Over de feitelijke effectiviteit van nascholingsprogramma's bestaat in de literatuur veel minder eenstemmigheid, dan over de noodzaak nascholing goed af te stemmen op adequate toetsingsgegevens. Tot in het midden van de jaren zeventig lag de nadruk op onderzoek naar de kwaliteit van de gehanteerde programma's en op aspecten van het leerproces (Rosinski, 1987). Volgens sommige auteurs zou het effect van specifieke nascholingsactiviteiten nauwelijks kunnen worden vastgesteld vanwege het grote aantal informatiebronnen dat artsen

beschikbaar hebben (Goldfinger, 1982). In deze opvattingen is een zekere kentering waar te nemen. Cohen & Davis (1986) onderscheiden, in navolging van Dixon (1978) vier niveaus waarop de evaluatie van nascholingsprogramma's zich kan richten: tevredenheid van de cursist met het programma; competentie-veranderingen (veranderingen in kennis, vaardigheden of attituden); veranderingen van gedrag in de praktijksituatie; en effect in termen van toegenomen gezondheid of welzijn van patiënten (patient outcome). De vier niveaus staan met elkaar in een hiërarchisch verband, waarbij het laatste niveau, het outcome niveau, het meest cruciaal is. Het ultieme criterium om de effectiviteit van nascholingsactiviteiten te beoordelen, is immers de mate waarin de activiteiten bijdragen aan een verbeterde zorg (Levine et al, 1984-II). Dit niveau is tevens echter het moeilijkst te onderzoeken. Gegevens die berusten op gedegen onderzoek, zijn dan ook schaars. Veel effectmetingen beperken zich tot het bepalen van de satisfactie met het gevolgde programma of tot verandering in de competentie (Berg, 1979; Davis et al, 1992). Deze 'lagere' niveaus bepalen overigens mede de effectiviteit van de hogere niveaus. De satisfactie van de deelnemers aan nascholing is bij voorbeeld van invloed op wat ze van het onderwijsprogramma opsteken (Bloemer & Poiesz, 1987). Hierbij is uiteraard de kwaliteit van het onderwijs in de nascholing van belang (Escovitz & Davis, 1990).

Behalve het feit dat effectmetingen van nascholingsactiviteiten zich veelal niet richtten op praktijkgedrag of patient outcomes, vormt ook de *methodologische kwaliteit* van een aantal onderzoeken een probleem. De onderzoeksgegevens die beschikbaar zijn, vertonen vaak methodologische gebreken (Berg, 1979). Voor zover er gegevens werden verzameld, waren de conclusies uit verschillende onderzoeken nogal eens tegenstrijdig. De conclusie, op basis van een review van de nascholingevaluatie-literatuur over de periode 1960-1977 (66 studies) eind jaren zeventig, was dat nascholing wel toename van kennis teweegbrengt maar weinig effect heeft op het praktijkgedrag (Bertram & Brooks-Bertram, 1977). In de literatuur ontstond vervolgens een uitgebreide discussie over de vraag of nascholing al dan niet effectief was. Lloyd & Abrahamson (1979) publiceerden een review van artikelen, over de periode 1960-1977, met betrekking tot de effectiviteit van nascholing. De meeste artikelen rapporteerden een positief effect op competentie, maar methodologisch waren de meeste onderzoeken niet sterk genoeg om harde conclusies te kunnen trekken. Geconcludeerd werd dat meer onderzoek moest worden verricht naar de factoren die ertoe leiden dat sommige nascholingsprogramma's wel en andere niet effectief zijn. Evered & Williams (1980) concludeerden op basis van een review van artikelen eveneens dat de validiteit van onderzoeksgegevens naar de effectiviteit van medische nascholing nogal eens dubieus was. Tevens werd vastgesteld dat wel kennistoename te behalen valt met nascholingsactiviteiten, maar dat de kennis alleen bekijft door periodieke bekrachtiging ervan. Door Stein (1981) werden acht onderzoeksrapporten opgenomen in een review naar het effect van nascholing op patient outcome maten. De conclusie was dat wel gedragsverandering kon worden aangetoond, maar dat slechts in één studie patient outcome betrokken was (Stein, 1981). Een van de eerste studies naar effect op patient outcome werd uitgevoerd op het gebied van hypertensie (Sackett et al, 1979). Het betrof hier echter relatief gemakkelijk te definiëren en meten criteria (medicatie en bloeddruk). Sibley et al (1982) concludeerden dat nascholing wel tot leereffect leidt, maar niet tot gedragsverandering. Sanazaro (1983) concludeerde, op basis van een review van de literatuur over de relatie tussen nascholing en de bevordering van competentie (gedefinieerd als 'ability to provide good care') c.q. performance, dat nascholing mag worden beschouwd als een effectief middel om artsen op de hoogte te houden van de

nieuwste ontwikkelingen op hun vakgebied, maar dat het geen garantie vormt dat artsen hun gedrag aan nieuwe inzichten aanpassen.

Medio jaren tachtig verschenen er opnieuw enkele omvangrijke reviews. Haynes et al (1984) deden verslag van 248 artikelen over nascholingsactiviteiten, gericht op het vaststellen van de werkzaamheid van nascholing. De bruikbaarheid en wetenschappelijke geloofwaardigheid van de artikelen werd bepaald met behulp van tevoren vastgestelde methodologische criteria met betrekking tot het gehanteerde onderzoeksdesign, het soort en het doel van de interventie, de doelgroep, de aanwezigheid van objectieve evaluatie van de nascholing in termen van satisfactiebepalingen, bepaling van kennistoename, gedragsverandering en/of patient outcome, statistische significantie en onderwijskundig en klinisch belang van de bevindingen. In 13% van de in de review betrokken onderzoeken bleek men gebruik te hebben gemaakt van een onderzoeksdesign in de vorm van een randomized control trial; 38% van de onderzoeken omvatte ook toetsing van klinische performance. In slechts 7% van de onderzoeken waren patient outcome maten betrokken. Minder dan de helft van de onderzoeken voorzag in enige statistische interpretatie van de bevindingen. Slechts 7 onderzoeken (3%) voldeden aan alle vastgestelde methodologische criteria. De conclusie was dat onderzoek naar de effectiviteit van nascholing blijkbaar veel methodologische en logistieke problemen met zich meebrengt, maar dat wellicht nascholing op zichzelf al van voldoende belang is (zodat rechtvaardiging in de zin van toename van de kwaliteit van de patiëntenzorg niet nodig is, als dat al zou kunnen worden aangetoond). Verder werd geconcludeerd dat het blijkbaar wel mogelijk is door nascholing gedrag te veranderen. Het effect van nascholing op patient outcome maten was echter nauwelijks aantoonbaar (Haynes et al, 1984). Davis et al (1984) en Nona et al (1988) kwamen tot vergelijkbare conclusies.

Naast de methodologische feilen van het uitgevoerde evaluatie-onderzoek, wordt in het afgelopen decennium in toenemende mate ook de gebrekkige *onderwijskundige kwaliteit* van veel nascholingsprogramma's beschouwd als een verklarende factor voor de teleurstellende bevindingen wat betreft de effectiviteit van de programma's. Kwalitatief goed nascholings-onderwijs heeft een aantal kenmerken. Het onderwijs moet in de eerste plaats zijn gebaseerd op principes van volwasseneneducatie (Stein, 1981; Cohen & Davis, 1986; Schofield, 1987). Dat betekent dat het onderwijs praktijkrelevant moet zijn, niet uitsluitend frontaal uitgevoerd, probleemgeoriënteerd en gericht op de behoeften van de deelnemers (Berg, 1979). Daarnaast speelt ook de herhaling van de boodschap die de nascholing beoogt over te brengen, en de bekrachtiging ervan door gezaghebbende deskundigen een rol (Horder, 1986; Kosecoff et al, 1987; Greer, 1988). Veel artsen veranderen hun door jarenlange praktijkervaring opgebouwde gewoontes niet op basis van één studie waarin ander gedrag wordt aanbevolen (Evered & Williams, 1980; Goldfinger, 1982; Wergin et al, 1983; Ferguson et al, 1984; Escovitz & Davis, 1990). Het vergaren van kennis wordt in het proces van gewenste gedragsverandering in de praktijk overigens beschouwd als een wel noodzakelijk maar op zichzelf onvoldoende element om deze verandering tot stand te brengen. Andere noodzakelijke elementen zijn: motivatie tot veranderen en het beschikken over het vermogen verschil te zien tussen ideaal en feitelijk praktijkgedrag (Geertsma et al, 1982). Hierbij is ook de dosering van de informatie en de 'verpakking' van belang, zoals blijkt uit bijvoorbeeld de didactische stijl die bij de overdracht van informatie wordt gehanteerd en de mate van praktijkgerichtheid van de informatie (Kanouse & Jacobi, 1988). In een review van 50 methodologisch verantwoord opgezette studies, naar de effectiviteit van nascholing wat betreft performance en patient

outcome, concluderen Davis et al (1992) dat nascholing voor deze uitkomstmaten effectief kan zijn onder een aantal condities. Deze condities betreffen met name de praktische relevantie en mate van actieve betrokkenheid van de nageschoolde artsen in het onderwijsprogramma en de afstemming van het onderwijs op objectief aangetoonde behoeften (Davis et al, 1992).

Kortom, als het onderwijs erin slaagt in te spelen op de behoeften van degenen die het onderwijs volgen, als het de lerenden weet te motiveren en als het wordt gegeven in een vorm die het leren bevordert, is de kans het grootst dat men daadwerkelijk profijt heeft van het gegeven onderwijs (Grosswald, 1984). De criteria voor effectieve nascholingsprogramma's worden door Harden (1992) als volgt samengevat: convenience (gemakkelijke deelname); relevance (sluit aan op het dagelijks werk in de praktijk); individualization (geeft deelnemers inspraak en mogelijkheden programma aan behoeften deelnemers aan te passen); self assessment (bevordert dokters het eigen begrip te toetsen en lacunes te vullen); interest (wekt aandacht en bevordert actieve deelname); speculation (erkent controversiële en grijze gebieden in geneeskunde); en systematic (programma is op een logische, systematische wijze opgezet) (CRISISS).

Het inzicht in de samenhang tussen kwalitatief goed (nascholings)onderwijs en effectief onderwijs, plus het toegenomen inzicht in de factoren die deze kwaliteit beïnvloeden, lijkt gunstige consequenties te hebben gehad. Beaudry concludeert in 1989 na een meta-analyse van 41 nascholingseffectiviteitsstudies, dat onderzoek en onderwijs de kinderschoenen te boven lijken. De conclusie van de studie is dat (goede) nascholing een positief effect lijkt te hebben zowel op de kennis en performance van artsen, als op de gezondheidstoestand van de patiënten (Beaudry, 1989). De volgende vraag is dan natuurlijk welk verband wordt aangetroffen tussen de prestaties op (kennis)toetsen en praktijkgedrag van artsen. De meest recente empirische gegevens wijzen op een positief verband tussen beide (Ramsey et al, 1989; Norman, 1991; Lockyer, 1992).

1.2.5. Vormen van kennistoetsing

De laatste invalshoek van waaruit de literatuur wordt beschreven, betreft de in kennistoetsen te hanteren vraagvorm. Het gaat hier dus om de vraag op welke wijze kennis het beste kan worden getoetst.

Halverwege de jaren zeventig werd in de geneeskunde het gebruik van gesloten, schriftelijke vraagvormen in kennistoetsen geïntroduceerd. Dit werd gemotiveerd vanuit de gedachte dat met dergelijke vraagvormen het beste objectieve, betrouwbare toetsen waren samen te stellen. Hieronder werden toetsen verstaan die door de deelnemers onder gelijke omstandigheden worden gemaakt, waarbij de antwoorden alle op dezelfde wijze worden beoordeeld, zonder subjectieve bias of invloed van docenten (Anderson, 1976). Van deze toetsen werden betrouwbare scores verwacht, hetgeen inhoudt dat de deelnemers zodanig worden gerangordend dat de slechtsten laag en de besten hoog scoren. De rol van het toeval bij het bepalen van de scores zou geminimaliseerd worden omdat alle kandidaten op gelijke wijze zouden worden behandeld (Anderson, 1979).

De opkomst van de objectieve toetsing ging gepaard met de vraag naar de beste vraagvorm voor dergelijke toetsen (Harden, 1979). Verschillende onderzoekingen kwamen niet altijd met gelijklopende resultaten voor de onderscheiden typen meerkeuzevragen. Anderson (1976) onderscheidde de 1-uit-5 keuzevraag en de multiële juist-onjuistvorm. Het eerste type zou

volgens de auteur beter discrimineren maar is moeilijker samen te stellen. Bovendien is er inhoudelijk een minder groot gebied mee te bestrijken dan met juist-onjuist-vragen. Deze laatste vorm riskeert op haar beurt weer meer gokgedrag, hetgeen volgens de auteur substantieel kan worden voorkomen door het toevoegen van een vraagtekenoptie. Frisbie (1973, 1974) vergeleek de betrouwbaarheid en de (concurrerende) validiteit van toetsen bestaande uit meerkeuzevragen danwel juist-onjuist vragen. Ook de toetstijd werd in het onderzoek betrokken. De meerkeuzevorm bleek veel betrouwbaarder scores (KR-20) op te leveren dan de juist-onjuistvorm. In dezelfde tijd konden deelnemers echter wel veel meer juist-onjuist vragen invullen dan meerkeuzevragen. Beide vraagvormen leken wel dezelfde kennis te meten. Oosterhof & Glasnapp (1974) onderzochten eveneens de vergelijkbaarheid van meerkeuze- en juist-onjuist vraagvormen als gecorrigeerd werd voor de noodzakelijke invultijd van beide vormen. Het resultaat was dat bijna 2.5 maal zoveel juist-onjuist vragen nodig waren om zelfde betrouwbaarheid te bereiken als met meerkeuzevragen. Ook Green (1978) vergeleek de betrouwbaarheid en concurrerende validiteit van equivalente toetsen bestaande uit hetzij meerkeuzevragen, hetzij juist-onjuist-vragen. In dit onderzoek werden geen significante verschillen gevonden tussen beide vormen.

Dat toetsen bestaande uit meerkeuzevragen, betrouwbare resultaten kunnen opleveren, wordt in de literatuur niet betwijfeld. De validiteit van deze toetsen was echter niet onomstreden (Norcini et al, 1984; Neufeld & Norman, 1985). Veel docenten waren sceptisch ten aanzien van de gebruikswaarde van de gesloten schriftelijke vraagvorm. Dit bezwaar was vooral gebaseerd op ongeloof dat externe instanties in staat zouden zijn het kennisniveau adequaat te onderzoeken. Bovendien waren veel docenten ervan overtuigd dat het aankruisen van een antwoord niet het eindpunt zou kunnen en mogen zijn van een voorafgaand ingewikkeld denkproces (van Berkel & Bax, 1990). Meerkeuzevragen zouden te artificieel, te absoluut en te gemakkelijk gericht zijn op trivialiteiten (Pickering, 1979). Daarnaast werd het als een nadeel beschouwd dat niet alle onderwijsdoelen ermee te toetsen zijn (zoals creativiteit, of het formuleren van meningen). Het beroep van deze vragen op de leesvaardigheid is in het algemeen groot, de kans op raden van het juiste antwoord is aanwezig en de constructie van dergelijke vragen is moeilijk en tijdrovend (Moelands e.a., 1992).

Volgens McGuire (1987) zijn de meeste bezwaren tegen schriftelijke gesloten vraagvormen terug te voeren op een aantal wijd verbreide misverstanden. Uiteindelijk bestaan er slechts twee basisvormen voor schriftelijke kennistoetsen: de zogenaamde 'free response' of open en de zogenaamde 'fixed response' of gesloten vraagvorm. In het eerste geval moet degene die de toets aflegt de antwoorden zelf formuleren, in het tweede geval moet het juiste antwoord worden geselecteerd. Het eerste misverstand is volgens McGuire dat de eerste vorm een subjectieve toetsvorm oplevert, en de tweede vorm een objectieve. Het verschil in objectiviteit en subjectiviteit is daarentegen graadueel en wordt bepaald door de procedures die worden gehanteerd bij de constructie en bij het nakijken van de vragen. Een ander misverstand is dat de vraagvorm het competentie-aspect dat wordt gemeten zou bepalen. Met beide typen vragen kan echter zowel pure feitenkennis als verdergaande cognitieve vaardigheden worden gemeten (McGuire, 1987).

De scepsis ten aanzien van de bruikbaarheid en relevantie van meerkeuzevragen heeft ertoe geleid dat de aandacht van toetsconstructeurs zich verplaatste naar het ontwikkelen van alternatieve toetsvormen, zoals patient management problems (PMP's), essay-achtige schriftelijke toetsvormen en gestructureerde vaardighedentoetsing (van der Vleuten & Newble,

1994). Vaardigheden vormen een ander competentie-aspect dan kennis. Toepassing van gestructureerde vaardigheidentoetsing op grote schaal brengt echter veel logistieke problemen met zich en de betrouwbaarheid van de scores is nogal eens problematisch (Jansen et al, 1992, 1993; Tan e.a., 1993).

Onderzoek naar de betrouwbaarheid, validiteit en efficiëntie van verschillende alternatieve kennistoetsvormen (op het gebied van de interne geneeskunde), liet zien dat meerkeuzevragen op alle fronten betere resultaten opleverden dan PMP's (Norcini et al, 1984, 1985; Swanson et al, 1987). Bovendien werd er geen ander competentie-aspect mee getoetst. Met name het scoren leverde bij PMP's problemen op naarmate de complexiteit van de schriftelijke patiëntsimulatie groter was. Bovendien wezen in de loop van de jaren tachtig diverse onderzoeken er steeds duidelijker op dat prestaties op de ene simulatie zich slecht lieten vertalen naar een andere simulatie. Dit werd het probleem van de 'inhoudsspecificiteit' genoemd (Norman et al, 1985; van der Vleuten & Newble, 1994). De consequentie daarvan is dat zeer lange toetstijden (van meer dan een dag) nodig zijn om tot betrouwbare scores te komen.

Vergelijking van meerkeuzevragen met meer of minder gestructureerde open essayvragen, maakte duidelijk dat alle vormen betrouwbare en valide scores kunnen opleveren. Er werd geen bewijs gevonden dat de verschillende vraagvormen verschillende cognitieve vaardigheden toetsen (Norman et al, 1987).

Swanson et al (1991) signaleerden eveneens dat schriftelijke en computergebaseerde klinische simulaties veel psychometrische problemen opleveren. De scores worden zeer beïnvloed door de antwoordstijl van de kandidaten, waardoor te lange toetsen nodig zijn om stabiele scores te verkrijgen. De correlaties met meerkeuzetoetsen zijn bovendien hoog. Toetsen met (korte) open vragen bieden dezelfde mogelijkheden als toetsen met meerkeuzevragen. Essay-examens hebben het nadeel dat een weinig breed gebied aan de orde kan worden gesteld, zodat ze per saldo veel meer invultijd kosten (Swanson et al, 1991). Schuwirth e.a. (1993) vinden, in een vergelijkend onderzoek naar korte casustoetsen met open vragen en toetsen met juist-onjuist-vraagtekenvragen, geen aanwijzingen dat een bepaald vraagtype meer meetinformatie oplevert dan een willekeurig ander. Ook Norman et al (1991) concluderen dat toetsen waarin gebruik wordt gemaakt van open of gesloten vraagvormen elkaar in betrouwbaarheid of validiteit niet veel ontlopen. Gesloten vraagvormen vergen meer constructietijd dan open vormen, maar die tijd wordt, als het om grote aantallen gaat, weer 'goedgemaakt' bij de scoring.

De conclusie van de verschillende onderzoeken naar de beste vraagvorm is dat welke vorm ook wordt verkozen, de inhoudelijke eisen waaraan de toets moet voldoen, in elk geval voorop staan. Bij schriftelijke toetsing van kennis is de vorm ondergeschikt aan de inhoud (Norman et al, 1991; van der Vleuten, 1991).

Ondanks hun relatieve impopulariteit zijn schriftelijke kennistoetsen met meerkeuzevragen, na hun introductie nooit uit het medisch onderwijs weggeweest (Norcini et al, 1985). Vooral bij landelijke toetsen met grote aantallen kandidaten is vrijwel altijd gebruik gemaakt van gesloten schriftelijke vraagvormen (Lockie, 1990; Spike & Veith, 1990; Norman et al, 1991). Ook in onderwijsvernieuingsprocessen werd voor het toetsen van kennis gebruik gemaakt van meerkeuzevragen (van der Vleuten & Verwijnen, 1990).

In zijn algemeenheid kunnen de voordelen ervan als volgt worden samengevat (Haladyna & Downing, 1989; van Berkel & Bax, 1990; Moelands e.a., 1992).

- in relatief korte tijd kan, ook bij grote aantallen deelnemers, een breed domein worden getoetst;

- de dekking van de inhoud van het vakgebied is meestal beter dan met andere meetmethoden mogelijk is;
- formuleervaardigheden van kandidaten worden niet ten onrechte verward met inhoudelijke kennis;
- er is een hoge betrouwbaarheid van de toetsscores mogelijk;
- vragen kunnen gemakkelijk worden opgeslagen, gebruikt en opnieuw gebruikt, met name als er een geautomatiseerde itembank bestaat;
- een objectieve en snelle toetsscorebepaling is mogelijk;
- diagnostische subscores zijn gemakkelijk te bepalen;
- de meeste soorten inhoud zijn met deze vorm toetsbaar, incl. allerlei soort denken op hoger abstractieniveau;
- de vorm is bij uitstek geschikt om de kwaliteit van de toetsen te onderzoeken en te verbeteren.

De opkomst van de theorie over het ontstaan van (medische) expertise en de rol van kennis daarbij, betekende een herwaardering voor het belang en de toetsing van kennis. Het hield tevens in dat de discussie over verschillende typen gesloten vraagvormen weer oplaaide. Als argumenten tegen het gebruik van de juist-onjuist vraagvorm worden opgevoerd dat de betrouwbaarheid van toetsen die bestaan uit dergelijke vragen nogal eens in het geding is; dat de vorm lastig te construeren is (zeker in de geneeskunde is iets maar heel zelden absoluut waar of onwaar); en dat het presenteren van onware stellingen zich slecht verdraagt met de gangbare onderwijssituatie, waarin een docent de waarheid overdraagt en geen onwaarheden (Moelands e.a., 1992). Ebel (1982) daarentegen beveelt juist-onjuist vragen aan, vanwege hun geschiktheid om kennisproposities te toetsen die de elementen weergeven van de structuurrelaties tussen concepten zoals die in het geheugen zijn opgeslagen. De betrouwbaarheidsproblemen van juist-onjuist vragen zijn volgens Ebel overschat.

De inhoudelijke keuze van Ebel voor juist-onjuist vragen werd later ondersteund, waarbij werd geadviseerd goed te letten op het niveau en de structuur van de inhoud waarop vragen zich moeten richten. Dan blijken juist-onjuist vragen zich heel goed te lenen om de beheersing van concepten te toetsen, zoals de 'clinical features' van ziektebeelden waartussen men wil discrimineren (Richardson, 1992). Het gebruik van schriftelijke casussimulaties heeft daarbij als voordeel dat ziekte- en patiëntgebonden factoren onder controle kunnen worden gehouden, omdat alle betrokkenen dezelfde casus maken (Jones et al, 1990). In de moderne vraagvormen is dan ook een trend waarneembaar naar gesloten vormen die patiëntgeoriënteerd zijn en realistische praktijksituaties weerspiegelen (Swanson & Case, 1992).

1.2.6. Conclusie

De literatuur over kennistoetsing in de huisartsgeneeskundige nascholing maakt een aantal zaken duidelijk. Systematisch verzamelde toetsingsgegevens worden van belang geacht omdat ze zowel op collectief als op individueel niveau bijdragen aan een objectiever inzicht in lacunes en sterke kanten van de huisarts (Grol & Zwaard, 1990). Kennistoetsen voor huisartsen moeten een kennisbegrip weerspiegelen dat recht doet aan het abstractieniveau dat de kennis van ervaren huisartsen kenmerkt. Dat abstractieniveau heeft niet zozeer een logisch, hypothetico-deductief karakter, maar wordt eerder gekenmerkt door patiëntgebonden probleemrepresentaties (Schmidt et al, 1990; Hobus, 1994). De wijze waarop kennis wordt getoetst, zou hierbij moeten aansluiten. Dat betekent dat in de toetsen de werkelijkheid van

de dagelijkse praktijkbeoefening zo dicht mogelijk moet worden benaderd (Norman, 1989; Metz, 1993; Schuwirth e.a., 1993; van der Vleuten & Newble, 1994). De toetsen moeten dus geloofwaardig zijn en relevante medische aspecten in de context van patiëntenproblematiek betreffen (Norcini & Dawson-Saunders, 1994).

De literatuur geeft dus een helder overzicht van de inhoudelijke eisen die aan relevante huisartsgeneeskundige kennistoetsen moeten worden gesteld. De literatuur laat echter ook zien dat over een aantal aspecten geen duidelijkheid bestaat. Hoe zien kennistoetsen die voldoen aan de eisen die door de theorie worden gesteld, er in de praktijk uit? In hoeverre kunnen kennistoetsen voldoen aan de gewenste gebruiksfuncties? Is kennistoetsing in de context van de praktijk van de Nederlandse huisartsgeneeskundige nascholing uitvoerbaar en acceptabel? Bestaat er samenhang tussen het kennisniveau van huisartsen en hun achtergrond- en praktijkkenmerken, en levert dit aanknopingspunten op voor de inrichting van de nascholing? Is het mogelijk met inhoudelijk verantwoorde kennistoetsen effectiviteit van nascholing aan te tonen? Bestaat er, gegeven de inhoud van de kennistoets, voorkeur voor een bepaalde (gesloten, schriftelijke) vraagvorm?

Deze vragen vormden de aanleiding tot het onderzoek dat in dit proefschrift wordt beschreven.

1.3. PROBLEEMSTELLING EN DOELEN ONDERZOEK

1.3.1. Probleemstelling

Het huisartsgeneeskundige kwaliteitsbeleid impliceert dat huisartsen gedurende hun professionele bestaan op gezette tijden blijf moeten geven van hun kennen en kunnen. Van toetsgegevens wordt verwacht dat ze objectieve indicaties voor bestaande lacunes aangeven waarop de nascholing, zowel op populatieniveau als op individueel niveau, zich vervolgens kan richten (Grol & Zwaard, 1990; LHV, 1993). De gedachte hierbij is dat nascholing gebaseerd moet zijn op objectief aangetoonde behoeften (in de zin van lacunes) van huisartsen. Daarnaast bestaat behoefte aan inzicht in de daadwerkelijke effectiviteit van goed opgezette en uitgevoerde nascholingsprogramma's.

Samengevat wordt een aantal *gebruiksfuncties* van kennistoetsgegevens onderscheiden:

a. *screening* van het algemene kennisniveau van huisartsen.

De screeningsfunctie die aan toetsgegevens wordt toegekend, impliceert dat de gegevens moeten bijdragen aan inzicht in eventueel bestaande verschillen in kennisniveau tussen diverse segmenten van de huisartspopulatie. Dit moet, op groepsniveau, leiden tot een gerichtere planning en sturing van de nascholing. De screeningsfunctie houdt in dat (periodiek) het niveau van kennis van ervaren huisartsen over het gehele huisartsgeneeskundige kennisdomein in kaart moet worden gebracht.

b. *educatief*, c.q. *diagnostisch hulpmiddel* voor individuele huisartsen.

De educatieve of diagnostische functie impliceert dat kennistoetsgegevens moeten bijdragen aan het inzicht van individuele huisartsen in hun eigen sterke en zwakke kanten in de kennis die van hen wordt verwacht.

c. *selectie* van huisartsen in het kader van (her)registratie.

In het kader van de problematiek rondom de (her)registratie van huisartsen bestaat behoefte aan objectieve gegevens op individueel niveau. Hierop kan men beslissingen over het al dan

niet voldoen aan een bepaalde norm baseren. De vraag hierbij is welke waarde aan kennistoetsgegevens voor dit doel kan worden gehecht.

d. *evaluatie* van onderwijsprogramma's.

Deze laatste gebruiksfunctie betreft de bruikbaarheid van kennistoetsgegevens om de effectiviteit van nascholingsprogramma's vast te stellen.

Het ontbreekt in de huisartsgeneeskunde vooralsnog aan een goed toepasbaar instrumentarium voor kennistoetsing, waarmee valide en betrouwbare gegevens voor bovenstaande functies systematisch kunnen worden verzameld. Daarbij gaat het zowel om instrumenten die gegevens opleveren over de algemene kennis van huisartsen, als om instrumenten die specifiek zijn gericht op het aantonen van effecten van gerichte (na)scholing. In dit onderzoek wordt gepoogd in dit gemis te voorzien door op het gebied van de huisartsgeneeskundige kennis toetsinstrumenten voor de beoogde gebruiksfuncties (verder) te ontwikkelen en op hun waarde en toepasbaarheid te onderzoeken. Daarbij is het streven instrumenten te ontwikkelen waarin een kennisbegrip wordt getoetst dat spoort met de theorie over het ontstaan van medische expertise.

Een vooronderzoek had duidelijk gemaakt dat algemene huisartsgeneeskundige kennistoetsen, waarin het gehele domein van relevante kennis steekproefsgewijs aan bod komt, geen werkbare basis zouden vormen om uitspraken te kunnen doen over de effectiviteit van gerichte nascholingsprogramma's (Düsmen, 1991). Om toch de verschillende gebruiksfuncties die ten aanzien van huisartsgeneeskundige kennistoetsing worden onderscheiden, in het onderzoek te betrekken, zijn twee vormen van kennistoetsing onderscheiden.

I Algemene, het huisartsgeneeskundig kennisdomein dekkende toetsen

Deze vorm is erop gericht algemene informatie op te leveren over de kennis van huisartsen over het gehele huisartsgeneeskundige vakgebied teneinde verschillen in kennis tussen huisartsen op het spoor te komen. De toetsen moeten, op groepsniveau, gebruikt kunnen worden bij de screening van het gemiddelde kennisniveau van huisartsen. Dit impliceert in de eerste plaats dat het huisartsgeneeskundig kennisdomein goed kan worden omschreven. Daarnaast betekent het dat het kennisbegrip dat wordt getoetst, zich inhoudelijk richt op de meest wezenlijke, praktijkrelevante elementen die moeten worden beheerst om huisartsgeneeskundige problemen adequaat te kunnen oplossen.

Naast informatie op geaggregeerd groepsniveau kunnen dergelijke toetsen een educatieve functie hebben, doordat informatie wordt geboden over individuele kennislacunes. Deze kennistoetsvorm moet tevens uitsluitsel geven over de potentiële selectieve functie: ook voor deze functie geldt in beginsel dat de toets het gehele kennisdomein moet dekken.

II Onderwerpgebonden kennistoetsen

De tweede vorm van kennistoetsing is er vooral op gericht informatie op te leveren over de kennis van huisartsen over bepaalde onderwerpen. Een belangrijke gebruiksfunctie van deze toetsvorm is de mogelijkheid om er effecten (in termen van kennistoename) van nascholingsactiviteiten mee te meten. Daarnaast kan men deze toetsvorm toepassen voor educatieve of diagnostische functies.

Het belangrijkste verschil tussen de twee kennistoetsvormen betreft de *toetsinhoud*, dat wil zeggen de inhoud van het gebied waarop de toets betrekking heeft en waarover de scores iets moeten zeggen. Op het niveau van de algemene kennistoets gaat het om de dekking van het

totale domein. De toetsscores staan niet in relatie met gespecificeerde onderwijsprogramma's. Op het niveau van de onderwerpgebonden toetsen gaat het nadrukkelijk wel om directe, gespecificeerde verbanden met de inhoud van (nascholings-)onderwijsprogramma's.

In het volgende schema (tabel 1.1) worden in een model de gebruiksfuncties voor de verschillende vormen van huisartsgeneeskundige kennistoetsen samengevat. Het model sluit aan op algemene modellen voor toepassingsdoelen van toetsgegevens (Millman & Greene, 1989; Eggen & Sanders, 1993).

Tabel 1.1. Gebruiksfuncties van huisartsgeneeskundige kennistoetsvormen.

Gebruiksfuncties	Vormen van kennistoetsen			
	Algemene kennistoets		Onderwerpgebonden toetsen	
	Groepsniveau	Individueel	Groepsniveau	Individueel
Screening	ondersteuning landelijke of regionale planning nascholing		keuze nascholingsprogramma door nascholingsgroepen	
Educatie		informatie over algemeen kennisniveau		informatie over kennisniveau over bepaalde onderwerpen
Selectie		voldoen aan bepaalde norm		
Evaluatie			bepalen effectiviteit nascholing	bepalen verandering in kennisniveau bij nascholing

1.3.2. Onderzoeksdoelen

Uit het onderzoek moet duidelijk worden op welke wijze verantwoorde instrumenten kunnen worden ontwikkeld voor het meten van huisartsgeneeskundige kennis, en wat de waarde en toepasbaarheid van de instrumenten is voor de verschillende gebruiksfuncties. Dat wil zeggen, voor het opsporen van kennislacunes op individueel en groepsniveau en het aangeven van de effectiviteit van nascholingsprogramma's.

Voor beide onderscheiden toetsvormen werden toetsen ontwikkeld. De volgende stap was het bepalen van de kwaliteit van de ontwikkelde toetsen als meetinstrumenten. Dat betekende dat de validiteit, betrouwbaarheid en praktische toepasbaarheid van de toetsen moest worden bepaald (Newble, 1992).

Het onderzoek beperkte zich niet tot het ontwikkelen van instrumenten. De instrumenten zijn ook daadwerkelijk in de praktijk toegepast om enkele gebruiksfuncties te kunnen toetsen. Dat betekent dat er een algemene kennistoets is ontwikkeld die vervolgens als screeningstoets is voorgelegd aan ervaren huisartsen, met het doel te achterhalen of bepaalde achtergrondkenmerken van huisartsen samenhang vertonen met de gemiddelde kennisscore van die huisartsen.

Daarnaast zijn enkele exemplarische onderwerpgebonden kennistoetsen ontwikkeld en afgenomen bij ervaren huisartsen om te achterhalen in hoeverre de toetsen in staat waren verandering in kennisniveau te laten zien in samenhang met het volgen van gerichte nascholing.

Wat betreft de vraagvorm in de toetsen is mede op grond van de brede toepasbaarheid die de instrumenten moesten hebben, gekozen voor een gesloten schriftelijke vorm. Omdat de literatuur duidelijk aantoonde dat de inhoud belangrijker is dan de vorm, is in het gehele onderzoek naar de inhoud van de toetsen de meeste aandacht uitgegaan. De literatuur maakte duidelijk dat wat betreft de vorm kon worden gekozen voor de meest handige, economische, logistiek haalbare en de minst tijdrovende. Om logistieke redenen werd in het onderzoek de juist-onjuist vraagvorm toegepast. Omdat uit de literatuur echter geen ondubbelzinnige conclusie was te trekken over de meest wenselijke specifieke gesloten vraagvorm, is dit aspect in een deel-onderzoek betrokken. Het doel daarvan was te achterhalen of, als de inhoud van de toetsing vast stond, nog voorkeur bestond voor een andere vorm dan voor de juist-onjuistvorm.

De *doelen* van het onderzoek worden als volgt samengevat:

- I. ontwikkeling van een praktisch toepasbaar instrument waarmee valide en betrouwbaar de algemene kennis van huisartsen in kaart kan worden gebracht;
- II. bepaling van de samenhang van verschillende professionele kenmerken met de algemene kennis van huisartsen;
- III. (exemplarische) ontwikkeling van onderwerpgebonden kennistoetsen die praktisch toepasbaar zijn en waarmee valide en betrouwbaar kennisverandering in samenhang met gerichte nascholingsprogramma's kan worden gemeten;
- IV. bepaling van effecten van de vraagvorm op de huisartsgeneeskundige kennistoetsing.

1.4. ONDERZOEKSVRAGEN

Op basis van de algemene doelen van het onderzoek, zijn onderzoeksvragen geformuleerd.

1.4.1. Ontwikkeling algemene huisartsgeneeskundige kennistoets

Doelstelling I: De ontwikkeling van een praktisch toepasbaar instrument waarmee valide en betrouwbaar de algemene kennis van huisartsen in kaart kan worden gebracht.

Uit de gebruiksfuncties van het instrument dat de algemene huisartsgeneeskundige kennis in kaart moet brengen, en uit de literatuur over het kennisbegrip dat in dergelijke toetsen aan de orde zou moeten komen, volgen inhoudelijke eisen waaraan het instrument moet voldoen. Het gaat hier immers om kennis die relevant moet zijn voor de huisartsgeneeskundige praktijkuitoefening. Daarnaast gelden psychometrische eisen. Toetsprocedures leiden tot scores: getallen die tot uitdrukking brengen in welke mate de getoetste personen beschikken over het kenmerk dat wordt getoetst. Het onderzoek naar de psychometrische kwaliteit van toetsen die deze kennis beogen te meten, moet aannemelijk maken dat de scores die op de toetsen worden behaald, een valide en betrouwbare weergave vormen van de mate waarin deelnemers aan de toetsen over deze kennis beschikken. Om te kunnen spreken van een goed

toepasbaar instrument, moet het daarnaast voor de doelgroep, namelijk ervaren huisartsen, ook acceptabel zijn. Tenslotte worden eisen gesteld aan de uitvoering en organisatie van de toetsing.

Dit leidt tot de volgende onderzoeksvragen.

1.1. Wat is de psychometrische kwaliteit van een algemene huisartsgeneeskundige kennistoets in termen van validiteit?

Om de screeningsfunctie te kunnen waarmaken, moeten de toetsresultaten inzicht geven in zowel relatief sterke als zwakke kanten in het kennisniveau van huisartsen. Dat betekent dat de toets als geheel een goede weergave moet vormen van het totale huisartsgeneeskundige kennisdomein. Om de kennistoets als diagnostisch/educatief hulpmiddel voor de individuele huisarts optimaal te kunnen gebruiken, dient de toets te bestaan uit elementen, waarvan de relevantie door de betrokkenen wordt onderkend. Over deze elementen moet bovendien zinvolle feedback worden gegeven. Met andere woorden, de gebruiksfuncties van de toets hebben consequenties voor de inhoudelijke specificering van de toets. Dit geldt zowel voor de reikwijdte, dat wil zeggen het domein dat de toets moet bestrijken, als voor de inhoud en detaillering van de toets (Millman & Greene, 1989).

De validiteit van een meetinstrument betreft de mate waarin de empirische bevindingen steun geven aan de veronderstelling dat het instrument een goede representatie vormt van het theoretische concept dat eraan ten grondslag ligt (Kane, 1982). Met andere woorden: een toets is valide als ermee wordt gemeten, wat men ermee beoogt te meten (Suen, 1990). Hiermee wordt een verband gelegd tussen het begrip validiteit en het gewenste gebruik van een toets, c.q. de gevolgtrekkingen die op basis van de scores mogen worden gemaakt (Messick, 1988). Validiteit wordt dan ook wel beschouwd als een oordeel, dat is gebaseerd op empirisch bewijs en theoretische principes, over de waarde van een toets voor bepaalde gebruiksfuncties. Naarmate met een toets beter aan de beoogde gebruiksfuncties tegemoet kan worden gekomen, is ze meer valide (Drenth, 1975). Een toets kan overigens voor het ene gebruik of de ene beslissing wel, en voor een ander gebruik of een andere beslissing niet valide zijn (van Berkel, 1984; Drenth & Sijsma, 1990).

De vragen die in het kader van de validering van de algemene huisartsgeneeskundige kennistoets relevant zijn, hebben in de eerste plaats betrekking op de criteria waaraan een huisartsgeneeskundige kennistoets moet voldoen om aan de gestelde gebruiksfuncties tegemoet te kunnen komen. Vervolgens wordt aandacht besteed aan de mate waarin een inhoudelijk goed onderbouwde kennistoets in staat is valide meetresultaten op te leveren.

1.2. Wat is de psychometrische kwaliteit van een algemene huisartsgeneeskundige kennistoets in termen van betrouwbaarheid?

Het begrip betrouwbaarheid van toetsscores betreft de mate waarin bij herhaling van de procedure hetzelfde resultaat wordt verkregen. Betrouwbaarheid wordt beschouwd als een noodzakelijke, maar op zich niet voldoende, voorwaarde om van validiteit van toetsgegevens te kunnen spreken (Thorndike, 1982; Feldt & Brennan, 1989).

De vraag naar de betrouwbaarheid is een vraag naar de accuratesse van de scores. In hoeverre komen de gevonden toetsscores overeen met de zogenaamde 'ware scores': de scores die uitsluitend worden bepaald door de prestatie van de deelnemers en niet door allerlei toevallige omstandigheden die de toetsafname beïnvloeden. Omdat toevallige omstandigheden bij elke meting een rol spelen, moet bij de interpretatie van toetsscores altijd met een bepaalde onbetrouwbaarheid rekening worden gehouden (Crocker & Algina, 1986). De mate van

onbetrouwbaarheid bepaalt onder meer de bruikbaarheid van een toetsinstrument. Naarmate de scores meer door het toeval worden bepaald, kan er minder betekenis aan worden gehecht (Neufeld & Norman, 1985).

1.3. Wat is de praktische toepasbaarheid in termen van acceptabiliteit en uitvoerbaarheid, van een algemene kennistoets voor huisartsen?

Inzicht in het oordeel van huisartsen over de inhoud en de relevantie van de toets als geheel en over de relevantie van de aparte toetsvragen, alsmede inzicht in hun opvattingen over de toepasbaarheid van de toets voor nascholingsdoelen is van belang om de acceptabiliteit van de ontwikkelde toetsmethode voor huisartsen te kunnen inschatten.

Daarnaast worden logistieke randvoorwaarden onderscheiden die de uitvoerbaarheid van de kennistoetsmethode bij gevestigde huisartsen betreffen. Verwacht mag worden dat de toepasbaarheid van de methode wordt beïnvloed door de benodigde inzet van mensen en middelen om toetsen te kunnen samenstellen, afnemen en verwerken, en om de toetsresultaten te kunnen terugrapporteren.

Er moet, met andere woorden, sprake zijn van een goed evenwicht tussen de kosten en de baten van de kennistoetsing.

1.4.2. Samenhang kennisniveau en achtergrondkenmerken huisartsen

Doelstelling II: Bepaling van de samenhang van verschillende professionele kenmerken met de algemene kennis van huisartsen.

Het tweede doel van het onderzoek naar de bruikbaarheid van kennistoetsing in de deskundigheidsbevordering van huisartsen, heeft betrekking op het achterhalen van bestaande verschillen in kennis tussen huisartsen, en op de factoren die eventueel bestaande verschillen kunnen verklaren.

De onderzoeksvragen in deze paragraaf richten zich op het aantonen van eventuele verschillen in kennisniveau tussen groepen huisartsen met verschillende achtergrondkenmerken. Dit leidt tot de volgende deelvragen.

II.1. Wat is het kennisniveau van verschillende groepen huisartsen op de algemene kennistoets als geheel en wat is het niveau op onderdelen van de kennistoets?

Deze onderzoeksvraag betreft het kennisniveau van huisartsen zoals dat wordt gemeten op de toets als geheel en op onderdelen van de toets. Het betreft een beschrijving van het kennisniveau van huisartsen, plus de variatie die daarbij wordt aangetroffen. In het onderzoek zal tevens worden betrokken hoe het kennisniveau van huisartsen zich verhoudt tot het kennisniveau van co-assistenten huisartsgeneeskunde en van huisartsen-in-opleiding.

Over de mate van aan te treffen verschillen in algemeen kennisniveau tussen huisartsen kunnen op basis van de literatuur geen expliciete verwachtingen worden geformuleerd. Wel is de groep huisartsen als geheel heterogener van samenstelling dan de verschillende groepen huisartsen-in-opleiding. Dat betekent dat wordt verwacht dat ook het gemiddelde kennisniveau van ervaren huisartsen meer uiteen zal lopen dan dat van aanstaande huisartsen. Behalve de vraag of over de hele linie kennisverschillen zijn waar te nemen tussen huisartsen en aanstaande huisartsen, of tussen huisartsen onderling, is ook de vraag van belang of specifieke hiaten in kennisniveau worden aangetroffen.

II.2. Hoe hangt het kennisniveau van huisartsen samen met bepaalde persoonlijke of professionele achtergrondkenmerken van deze huisartsen?

Het gaat er hierbij om na te gaan of eventueel aangetroffen verschillen in kennisniveau tussen huisartsen vanuit verschillende achtergrondkenmerken kunnen worden verklaard.

Op basis van de gegevens uit de literatuur bestond er een aantal verwachtingen over samenhangen tussen professionele kenmerken van huisartsen en hun kennisniveau. Met name leeftijd, ervaring, en mate van geïnvolveerdheid in onderwijs of nascholing lijken samenhang te vertonen met het huisartsgeneeskundige kennisniveau.

II.3. In hoeverre hangt het oordeel dat huisartsen zelf aangeven over hun specifieke expertise samen met hun kennisniveau?

Het in de literatuur aangegeven verschil tussen subjectieve en objectieve leerbehoeften en het effect daarvan op de effectiviteit van nascholingsprogramma's, maakt de vraag interessant in hoeverre huisartsen hun eigen deskundigheid realistisch inschatten. Met andere woorden, zijn ze ook goed in de onderwerpen waarover ze denken veel te weten?

1.4.3. Ontwikkeling kennistoets ter bepaling effectiviteit nascholing

Doelstelling III: (Exemplarische) ontwikkeling van onderwerpgebonden kennistoetsen die praktisch toepasbaar zijn en waarmee valide en betrouwbaar kennisverandering in samenhang met gerichte nascholing kan worden gemeten.

Dit deel van het onderzoek betreft de vraag in hoeverre het mogelijk is kennistoetsen te ontwikkelen die in staat zijn kennisveranderingen te meten onder invloed van een onderwerpgebonden, specifiek (nascholings-)programma. Er wordt nagegaan of de ontwikkelde instrumenten gevoelig genoeg zijn om verandering in kennisniveau te meten bij huisartsen die speciale onderwijsprogramma's hebben gevolgd.

De onderzoeksvragen behorend bij de onderwerpgebonden kennistoetsen, richten zich op de mogelijkheid dergelijke toetsen samen te stellen, op de psychometrische kwaliteiten ervan en op de gebruiksmogelijkheden van dergelijke toetsen om kennisontwikkeling in samenhang met nascholing te achterhalen.

III.1. Wat is de validiteit van onderwerpgebonden kennistoetsen die bij specifieke nascholingsprogramma's worden ontwikkeld?

In deze onderzoeksvraag over de validiteit van onderwerpgebonden kennistoetsen gaat het om de vraag of er kennistoetsen kunnen worden ontwikkeld die representatief zijn voor de meest wezenlijke kenniselementen in nascholingsprogramma's rondom een specifiek thema. De toets moet ook gevoelig genoeg zijn om verschillen in kennisniveau voorafgaand en volgend op nascholing te meten.

III.2. Wat is de betrouwbaarheid van onderwerpgebonden kennistoetsen die bij specifieke nascholingsprogramma's worden ontwikkeld?

Analoog aan de vraag naar de betrouwbaarheid van algemene kennistoetsen, is ook de vraag van belang naar de betrouwbaarheid van specifieke toetsen. Het gaat daarbij zowel om de betrouwbaarheid van individuele toetsscores als om de betrouwbaarheid van groepsgegevens.

III.3. Wat is de praktische toepasbaarheid in termen van acceptabiliteit en uitvoerbaarheid, van voor de nascholing bestemde, onderwerpggebonden kennistoetsen?

Ook voor de specifieke toetsen geldt dat uitspraken over de toepasbaarheid ervan voor nascholingsdoelen pas kunnen worden gedaan als daarbij ook de opinie van de deelnemers aan dergelijke toetsen is betrokken en als informatie beschikbaar is over de uitvoerbaarheid van de toetsingsprocedures: de logistieke kanten van de toetsontwikkeling, afnameprocedures en verwerkings- en feedbackprocedures.

1.4.4. Invloed vraagvorm op huisartsgeneeskundige kennistoetsing

Doelstelling IV: Bepaling van effecten van de vraagvorm op de huisartsgeneeskundige kennistoetsing.

Het laatste deel van het onderzoek betreft de invloed van de vraagvorm op de huisartsgeneeskundige kennistoetsing. Op grond van de literatuur is geen eensluidend antwoord te verkrijgen op de vraag naar de beste vraagvorm. Wel is duidelijk dat de inhoudelijke eisen waaraan goede toetsvragen moeten voldoen, belangrijker zijn dan de uiteindelijke vorm waarvoor wordt gekozen (van der Vleuten & Newble, 1994). De vraagvorm wordt in het onderzoek betrokken om argumenten te verzamelen die kunnen helpen bij de keuze voor de beste vorm. Specifieke onderzoeksvragen betreffen de invloed van toepassing van een andere schriftelijke vraagvorm dan de juist-onjuist-vraagtekenvorm, op de validiteit, de betrouwbaarheid en de praktische toepasbaarheid van huisartsgeneeskundige kennistoetsing.

IV.1. Welke invloed heeft de vraagvorm op de validiteit van de huisartsgeneeskundige kennistoetsen?

Deze vraag is erop gericht na te gaan of variatie in (schriftelijke) vraagvorm van invloed is op de validiteit van de toetsen. Daarbij gaat het er vooral om te achterhalen in hoeverre met verschillende vraagvormen dezelfde inhoudelijke kenniselementen worden getoetst.

IV.2. Welke invloed heeft de vraagvorm op de betrouwbaarheid van de huisartsgeneeskundige kennistoetsen?

Deze vraag betreft de invloed van verschillende vraagvormen die in gesloten schriftelijke kennistoetsen gehanteerd worden, op de betrouwbaarheid van de scores. Bij deze vraag is bijvoorbeeld aan de orde of verandering van de vraagvorm van invloed is op het aantal vragen dat nodig is om een bepaalde toetsbetrouwbaarheid te bereiken.

IV.3. Welke invloed heeft de vraagvorm op de praktische toepasbaarheid van de huisartsgeneeskundige kennistoetsen?

Evenals in de overige onderzoeksdelen is ook in dit deel van het onderzoek de vraag naar de praktische toepasbaarheid in termen van logistieke uitvoerbaarheid en acceptabiliteit voor de praktiserende huisartsen van belang.

DEEL I (hoofdstuk 2, 3 en 4)

ONTWIKKELING ALGEMENE KENNISTOETS VOOR HUISARTSEN

In het eerste deel van het onderzoek dat in dit proefschrift wordt beschreven, wordt ingegaan op welke wijze en met welk resultaat de vraagstellingen behorend bij het eerste onderzoeksdoel zijn beantwoord. Deze luiden als volgt.

Doelstelling I:

De ontwikkeling van een praktisch toepasbaar instrument waarmee valide en betrouwbaar de algemene kennis van huisartsen in kaart kan worden gebracht.

Vraagstellingen:

- I.1. Wat is de psychometrische kwaliteit van de algemene huisartsgeneeskundige kennistoets in termen van validiteit?
- I.2. Wat is de psychometrische kwaliteit van de algemene huisartsgeneeskundige kennistoets in termen van betrouwbaarheid?
- I.3. Wat is de praktische toepasbaarheid in termen van acceptabiliteit en uitvoerbaarheid van een algemene kennistoets voor huisartsen?

In de volgende drie hoofdstukken worden achtereenvolgens de validiteit, de betrouwbaarheid en de praktische toepasbaarheid beschreven van een algemene kennistoets voor huisartsen. Bij de ontwikkeling van deze toets is voortgebouwd op de ervaringen die zijn opgedaan met de landelijke kennistoets die als voortgangstoets is ontwikkeld voor de huisartsopleiding. Hierna wordt beargumenteerd waarom deze toets in beginsel ook geschikt werd geacht als instrument voor het meten van kennis van ervaren huisartsen.

- landelijke kennistoetsing in de beroepsopleiding tot huisarts

Voor de toetsing van de kennis van huisartsen-in-opleiding is door het landelijk Samenwerkingsverband van de Universitaire Huisartsopleidingen (SVUH), een methode ontwikkeld die informatie over veranderingen in het kennisniveau van aanstaande huisartsen gedurende hun opleiding beoogt op te leveren (Pollemans e.a., 1991). Om de aanstaande huisartsen inzicht te bieden in hun kennisontwikkeling, én tevens om -op beleidsniveau- te kunnen rapporteren over de gemiddelde kennisontwikkeling gedurende de opleiding, werd gekozen voor het zogenaamde 'voortgangstoetsmodel' (van der Vleuten & Verwijnen, 1990). Sinds 1988 worden drie maal per jaar, op vaste tijdstippen gedurende de gehele huisartsopleiding, kennistoetsen afgenomen bij alle huisartsen-in-opleiding. Alle artsen die op dat moment de opleiding volgen maken een identieke toets. De inhoud van de toets is steeds afgestemd op het eindniveau van de huisartsopleiding. Dit eindniveau is afgeleid van het LHV-Basistakenpakket van de huisarts (Springer, 1983). De kennistoets richt zich op de kennis die

noodzakelijk is om aan de eisen die het Basistakenpakket stelt, te voldoen, dat wil zeggen op de kennis die moet worden beheerst om de taken te kunnen uitvoeren, die door de beroepsgroep zijn geaccepteerd als behorend tot de verantwoordelijkheid van de huisarts.

Per keer nemen aan de landelijke kennistoets ongeveer 500 huisartsen-in-opleiding deel. Daarnaast zijn ongeveer 100 huisartsen op de toets geabonneerd. Dit grote aantal deelnemers maakt het gebruik van een schriftelijke, gesloten *vraagvorm* noodzakelijk. De vragen zijn geformuleerd als stellingen over praktijkrelevante huisartsgeneeskundige casuïstiek. De vragen betreffen elementen van de diagnostiek, behandeling of epidemiologie van het probleem waarop de casus betrekking heeft (van Leeuwen e.a., 1992). Hiermee wordt beoogd hogere cognitieve vaardigheden te toetsen dan eenvoudige feitenkennis (van Leeuwen & van Hessen, 1990; van Leeuwen, in voorbereiding).

De kennistoets bestaat steeds uit een 70- tot 80-tal casus waarover stellingen moeten worden beantwoord. Dit vergt een toetstijd van ongeveer twee uur. Alle vragen zijn voorzien van gedetailleerde literatuurreferenties. De toetsvragen moeten worden beantwoord met 'juist', 'onjuist' of 'vraagteken'. De 'vraagtekenoptie' is om educatieve redenen ingevoerd: het geeft deelnemers aan de toets de gelegenheid te erkennen dat ze kennis ontberen. In overeenstemming met aanbevelingen uit de literatuur wordt het gebruik van het 'vraagteken' aangeraden als deelnemers aan de toets het antwoord op de vraag absoluut niet weten (Fleming, 1988). Toetsvragen worden opgesteld door stafleden van alle acht huisartsopleidingen. Een landelijke werkgroep van huisartsen en toetsdeskundigen (de Werkgroep Kennis van het Samenwerkingsverband) beoordeelt de aangeleverde vragen op vorm, inhoud en relevantie, en corrigeert de vragen in een aantal commentaarronden (Kramer & Pollemans, 1990).

De kennistoetsen worden onder examencondities afgenomen tijdens het instituutsonderwijs van de huisartsen-in-opleiding. De deelnemers hebben de gelegenheid om de toetsvragen schriftelijk te bekritiseren. De antwoordformulieren van de deelnemers worden centraal verwerkt. Per toetsvraag wordt het percentage goede antwoorden, het percentage foute antwoorden en het percentage vraagtekens berekend, alsmede itemtest-correlaties en discriminatie-indices. Deze gegevens moeten duidelijk maken in hoeverre de individuele toetsvragen consistent worden beantwoord door de deelnemers aan de toets (Bender, 1994). Toetsvragen die achteraf om inhoudelijke of technische redenen van twijfelachtige kwaliteit blijken te zijn, worden weggelaten uit de definitieve versie waarover het toetsresultaat wordt berekend. Het verwijderen van toetsvragen (items) uit de toets gebeurt onder verantwoordelijkheid van de toetsbeoordelingscommissie, die voor die gelegenheid wordt aangevuld met drie onafhankelijke huisartsen.

Een goed antwoord levert één punt op, voor een fout antwoord wordt een punt afgetrokken. Vraagtekens tellen niet mee. Om te corrigeren voor gokgedrag geldt als toetsresultaat de 'goed-min-foutscore', die varieert tussen -100 en +100. Deze score is uitgedrukt als percentage van de maximaal te behalen score. In de instructie die deelnemers aan de toetsen krijgen, is toegelicht dat in de analyses de foute antwoorden in mindering worden gebracht op de goede.

De toetsresultaten worden individueel teruggedirigeerd, waarbij alle deelnemers de antwoordsleutel krijgen en een naar onderwerp gespecificeerd overzicht van de eigen resultaten. Deze zijn voorzien van de vergelijkbare gegevens van de eigen opleidingsgroep en van de landelijke gegevens van alle deelnemers die, op het moment van de toetsafname,

in dezelfde fase van hun opleiding verkeerden. Aan de toetsresultaten worden geen consequenties verbonden in de zin van zakken of slagen. Het resultaat is uitsluitend bestemd voor feedback-doeleinden voor de huisarts-in-opleiding en voor het opleidingsinstituut zelf.

De kennistoets is bedoeld als voortgangstoets waarbij het niveau is gedefinieerd als het eindniveau van de opleiding. Bij de start van het onderzoek waarover in dit proefschrift wordt gerapporteerd, was alleen bekend dat in de tweejarige opleiding tot huisarts een gemiddelde stijging in algemeen kennisniveau (uitgedrukt in goed-min-foutscore op de totale kennistoets) werd gevonden van ongeveer 15%. Dit stijgingspercentage was tamelijk constant voor de verschillende toetsen die in de huisartsopleiding waren afgenomen (Pollemans & Tan, 1990; Pollemans e.a., 1991). Het stijgingspercentage is vergelijkbaar met de stijging in kennistoetscores die wordt gevonden bij voortgangstoetsen over eenzelfde periode in bijvoorbeeld fysiotherapie-opleidingen en in de eindfase van het universitaire medisch curriculum (Besseling e.a., 1987; van der Vleuten & Verwijnen, 1990; van Hessen et al, 1991).

- bruikbaarheid kennistoetsmethode in nascholing

De methode van kennistoetsing die in de huisartsopleiding is ontwikkeld, voldeed in elk geval aan een belangrijke inhoudelijke eis, waaraan een kennistoets moet voldoen om geschikt te zijn voor toepassing binnen de beroepsgroep. De toets was gericht op de kennis waarover een huisarts moet beschikken om het LHV-Basistakenpakket (Springer, 1983) naar behoren te kunnen uitvoeren (Pollemans & Tan, 1990).

Daarnaast vormde het feit dat de toetsvragen betrekking hadden op praktijkrelevante huisartsgeneeskundige casuïstiek waarbij specifiek wordt beoogd te vragen naar de meest wezenlijke elementen van diagnostiek, beleid of epidemiologie ten aanzien van de casus, een belangrijke ondersteuning voor de gedachte dat de toets in beginsel ook toepasbaar zou zijn voor de toetsing van de algemene kennis van ervaren huisartsen. Het instrument was bovendien goed hanteerbaar bij landelijk georganiseerde toetsing van grote aantallen deelnemers.

Op grond hiervan is de kennistoetsmethode zoals ontwikkeld door het landelijk Samenwerkingsverband, als uitgangspunt genomen voor de studie naar een verantwoorde toets gericht op de algemene kennis van huisartsen. Dat betekende onder meer dat dezelfde toets kon worden voorgelegd aan huisartsen-in-opleiding en aan ervaren huisartsen. Behalve praktisch voordeel bood dit ook de gelegenheid om de toetsresultaten van huisartsen in verschillende fasen van hun opleiding of hun beroeps carrière met elkaar te vergelijken.

In de literatuur wordt een reeks aspecten onderscheiden die bij de specificatie van toetsinstrumenten duidelijk moeten worden (Millman & Greene, 1989; Eggen & Sanders, 1993). Deze aspecten betreffen externe factoren en interne toetskenmerken. Tot de externe factoren worden gerekend de doelgroep bij wie de toets moet worden afgenomen, de beschikbare toetstijd en de wijze waarop de toets wordt afgenomen. Het doel en de doelgroep voor de huisartsgeneeskundige toetsing zijn beschreven in het eerste hoofdstuk van dit verslag. De keuze om aan te sluiten bij de toetsontwikkeling van het landelijk Samenwerkingsverband impliceerde dat ook de overige beschreven externe factoren vast lagen.

De inhoud van de toets wordt beschouwd als belangrijkste intern kenmerk (Millman & Greene, 1989). Bij de start van het onderzoek was het instrument van het Samenwerkingsverband op een aantal punten inhoudelijk nog niet voldoende uitgekristalliseerd. Er was bij

voorbeeld nog geen sprake van een vaste inhoudsstructuur voor de toets, zodanig dat de inhoud van de toets een goede afspiegeling vormde van het huisartsgeneeskundig kennisdomein. De onderwerpen waaruit opeenvolgende toetsen bestonden, werden aangeleverd door dezelfde stafleden van de huisartsinstituten die ook de toetsvragen maakten. Het enige criterium waaraan de onderwerpen moesten voldoen, was dat ze huisartsgeneeskundig relevant moesten zijn. In de loop van de tijd was echter gebleken dat de voorraad 'spontaan' aangedragen onderwerpen afnam en dat de onderwerpen, voor zover ze wel werden aangedragen, steeds meer op 'huisartsgeneeskundige randgebieden' betrekking hadden (Pollemans e.a., 1993). Ten gevolge van de vrije onderwerpkeuze was geen enkele kennistoets een adequate afspiegeling van het gehele huisartsgeneeskundige kennisdomein. Om haar voortgangstoetsfunctie in de huisartsopleiding te kunnen waarmaken, was het noodzakelijk dat de verschillende opeenvolgende toetsen inhoudelijk steeds een gelijksoortige representatie van het gehele kennisdomein zouden vormen. Dat betekende dat een blauwdruk voor de toets, die een dergelijke specificatie moest bieden, noodzakelijk werd (Pollemans & Tan, 1990). Als de toets ook moest worden toegepast om de kennis van ervaren huisartsen in kaart te brengen, was een dergelijke specificatie eveneens noodzakelijk.

Andere interne toetskenmerken die in de literatuur worden opgesomd, lagen met de keuze voor aansluiting bij de huisartsopleidingstoets wél vast. Dit betrof de vorm en de constructiewijze van de toetsvragen, het aantal toetsvragen, en de scoringsprocedures (Millman & Greene, 1989; Eggen & Sanders, 1993).

De eerste activiteit in het kader van dit deel van het onderzoek was het vaststellen van een relevante toetsblauwdruk. Het resultaat van deze activiteit wordt beschreven in hoofdstuk 2, waarin wordt ingegaan op de validiteit van de huisartsgeneeskundige kennistoets.

Over de betrouwbaarheid van de toetsgegevens was bij de start van het onderzoek nog weinig bekend. In het kader van het onderzoek naar de toepasbaarheid van de algemene kennistoets voor ervaren huisartsen, is daarom veel aandacht geschonken aan de betrouwbaarheid van de toetsgegevens. Het resultaat wordt beschreven in hoofdstuk 3. In hoofdstuk 4 tenslotte wordt ingegaan op de acceptabiliteit en praktische uitvoerbaarheid van algemene kennistoetsing voor ervaren huisartsen.

Hoofdstuk 2

VALIDITEIT ALGEMENE HUISARTSGENEESKUNDIGE KENNISTOETS

SAMENVATTING

De studie naar de validiteit van de algemene huisartsgeneeskundige kennistoets is gericht op het aannemelijk maken van de inhouds- en van de constructvaliditeit van de toets.

In de studie naar de inhoudsvaliditeit staat het onderbouwen en verantwoorden van de inhoudelijke richtlijnen (blauwdruk) voor het samenstellen van toetsen centraal. De blauwdruk moet ertoe leiden dat de toetsen evenwichtige afspiegelingen vormen van het huisartsgeneeskundig kennisdomein. De blauwdruk moet er bovendien toe leiden dat dit kennisdomein in de toets, vanuit verschillende perspectieven, adequaat wordt geoperationaliseerd. De blauwdruk is op basis van inhoudelijke criteria ontwikkeld, in een stapsgewijze consensusprocedure met experts op het terrein van de huisartsgeneeskundige scholing en toetsing. Als hoofdindeling is gekozen voor aansluiting bij de klachten- en aandoeningencomponenten van de International Classification of Primary Care (ICPC) (Lamberts & Wood, 1987), aangevuld met een huisartsgeneeskundig-theoretisch onderdeel. Daarnaast is voor vier nevenindelingen gekozen: leeftijdscategorieën van patiënten, aspecten van het consult, spoedeisende gevallen en chronische aandoeningen. De relatieve omvang van de verschillende blauwdrukcategorieën is vastgesteld op inhoudelijke gronden. In het onderzoek werd nagegaan of er meettechnische argumenten bestonden om wijzigingen aan te brengen in de omvang van de categorieën. Het bleek dat de feitelijk gehanteerde omvang van de inhoudsgebieden nauwelijks afweek van de vanuit betrouwbaarheidsoogpunt optimale omvang. Dit gegeven wordt beschouwd als een empirische ondersteuning van de inhoudsvaliditeit van de algemene huisartsgeneeskundige kennistoets.

Op basis van de blauwdruk is een kennistoets samengesteld en voorgelegd aan groepen aanstaande artsen en aanstaande huisartsen, en aan ervaren huisartsen in verschillende fasen van hun beroeps carrière. De gegevens die met deze toetsafname zijn verkregen, zijn onder meer gebruikt om de onderzoeksvragen te beantwoorden over de constructvaliditeit van de toets. Voor deze studie is aangesloten bij de cognitief-psychologische theorievorming over het ontstaan van (medische) expertise. Verwacht werd dat het kennisniveau het hoogst is aan het eind van de huisartsopleiding. Daarbij wordt ervan uitgegaan dat de ontwikkeling van kennis over huisartsgeneeskundig-inhoudelijke concepten, het meest geconcentreerd plaatsvindt tijdens de huisartsopleiding, en dat de kennisontwikkeling in de periode daarna vooral samenhangt met voor elke huisarts verschillende praktijkervaringen. Uit de resultaten bleek dat de ervaren huisartsen als groep gemiddeld hogere scores behaalden dan de groep huisartsen-in-opleiding, die op haar beurt weer hogere gemiddelde scores behaalde dan de co-assistenten huisartsgeneeskunde. De verschillen in gemiddelde toetsscore van deze groepen

waren statistisch significant. De ervaren huisartsen bereiken een gemiddelde toetsscore die ongeveer op het niveau lag dat door huisartsen-in-opleiding vlak voor het einde van de huisartsopleiding wordt bereikt. Deze bevindingen worden beschouwd als ondersteuning van de constructvaliditeit van de algemene huisartsgeneeskundige kennistoets.

2.1. INLEIDING

Validiteit gaat over de betekenis, de bruikbaarheid en de juistheid van de conclusies die uit toetsscores worden getrokken (APA-Standards, 1985; Messick, 1989; Kane, 1992). Om de validiteit van een toets te bepalen, worden de volgende drie benaderingen onderscheiden: inhoudsvaliditeit, criterium-gerelateerde validiteit en constructvaliditeit (Eggen & Sanders, 1993). Het betreft hier niet zo zeer aparte validiteitstypen, als wel verschillende benaderingen van validiteitsonderzoek (Suen, 1990).

Inhoudsvaliditeit verwijst naar de mate waarin de toets een goede afspiegeling vormt van het te toetsen domein. Het gaat hierbij zowel om de relevantie van de items in relatie tot het beoogde gebruik van de toets, als om de representativiteit van de items voor het te toetsen domein (Suen, 1990). Bij het onderzoek naar de inhoudsvaliditeit van een toets ligt de nadruk op de grenzen van het inhoudelijke domein dat met een instrument wordt bestreken, en niet zozeer op de toetsscores die worden behaald (Messick, 1975, 1988). In de andere validiteitsbenaderingen ligt de nadruk wel op de betekenis, de interpretatie die aan de scores kan worden gegeven. *Criteriumvaliditeit* betreft de voorspellende waarde van de toetsscores ten aanzien van een bepaalde externe (gelijktijdig gemeten of in de toekomst te bepalen) criterium-maat. Deze maat wordt beschouwd als een soort 'gouden standaard'. Als er geen goede criterium-maat bestaat, is criteriumvalidering van een instrument niet zinvol. Als er wel een goede criterium-maat is, betekent dat dat er al een valide instrument (maar mogelijk duurder, of complexer) bestaat voor het object van onderzoek. Het nieuwe instrument kan daarmee worden vergeleken (Suen, 1990). *Constructvaliditeit* betreft de mate waarin een instrument in staat is de theoretisch gefundeerde eigenschap te meten (Cronbach & Meehl, 1955). Bij het aannemelijk maken van constructvaliditeit gaat het erom of de toetsscores een goede indicator vormen van dit theoretische construct (Messick, 1975). Een goede theorie over het construct dat wordt gemeten, staat dan ook centraal in het onderzoek naar de constructvaliditeit van een toets (Suen, 1990). Constructvalidering is noodzakelijk om uitspraken te kunnen doen over de mate waarin aan toetsscores specifieke gevolgtrekkingen mogen worden verbonden (APA-Standards, 1985). Afhankelijk van het soort gevolgtrekkingen dat men wenst te maken, kan dezelfde test verschillende (construct-)validiteiten hebben (Swanson et al, 1987).

In de literatuur is uitgebreid gediscussieerd over de vermeende (meer)-waarde van de ene benadering boven de andere. De verschillende benaderingen sluiten elkaar echter niet uit, maar moeten worden beschouwd als elkaar aanvullend. Met betrekking tot toetsen die bestemd zijn voor onderwijs- en opleidingssituaties, geldt meer dan voor andere meetinstrumenten dat ze moeten bestaan uit concrete taken die van direct praktisch belang zijn voor de latere beroepsuitoefening (Ebel, 1983). Voor dergelijke toetsen is het aantonen van inhoudsvaliditeit, dat wil zeggen het aannemelijk maken dat de toetsinhoud een realistische afspiegeling vormt van de praktijk waarvoor het onderwijs opleidt, van cruciaal belang (Ebel, 1961, 1983). Ebel bepleit in dit verband het domein dat in de toets aan de orde komt, op basis van een goed

onderbouwde analyse zorgvuldig te verantwoorden. Andere auteurs benadrukken dat in de validiteitsdiscussie de interpretatiemogelijkheden van de toetsscores doorslaggevend moeten zijn en leggen daarmee het primaat bij het aantonen van constructvaliditeit (Cronbach, 1988, 1989). Dat impliceert dat op basis van een theorie relaties tussen variabelen moeten worden gespecificeerd die vervolgens empirisch moeten worden onderzocht (Cronbach & Meehl, 1955). Deze relaties kunnen ook inhoudelijk van aard zijn. De noodzaak van constructvalidering van toetsinstrumenten is onomstreden. Daarnaast is er echter de laatste jaren sprake van een hernieuwde belangstelling voor de inhoudsvaliditeit van toetsen die bestemd zijn voor gebruik in het onderwijs (Newble et al, 1994). Daarom is in de valideringsstudie van de algemene huisartsgeneeskundige kennistoets ervoor gekozen nadruk te leggen op de inhoudelijke onderbouwing van de toets.

Onderzoeksvraag 1.1.:

Wat is de psychometrische kwaliteit van een algemene huisartsgeneeskundige kennistoets in termen van validiteit?

Opzet validiteitsstudie:

- blauwdrukontwikkeling
- empirische inhoudsvalidering blauwdruk: multivariate generaliseerbaarheidsanalyse blauwdruk
- constructvalidering kennistoets: scoreverloop en expertise-ontwikkeling

2.2. ONDERZOEKSOPZET VALIDITEITSSTUDIE

Het valideringsonderzoek is gericht op de inhoudsvaliditeit en op de constructvaliditeit van de toets. Wat betreft het bepalen van de criteriumvaliditeit van de toets, wordt het in de literatuur gehuldigde standpunt gevolgd dat er voor kennis geen algemeen aanvaard criterium, een 'gouden standaard', bestaat (Kane, 1982, 1987). Daarom blijft deze validiteitsbenadering hier verder buiten beschouwing.

2.2.1. Bepaling inhoudsvaliditeit

De gebruiksdoelen van de algemene huisartsgeneeskundige kennistoets bepalen het inhoudelijk domein dat de toets moet bestrijken. Een voorwaarde voor inhoudsvaliditeit is dat de toets voldoende representatief is voor dit domein. Het huisartsgeneeskundig kennisdomein werd afgebakend op geleide van een aantal criteria. Op basis daarvan is een *blauwdruk*, dat wil zeggen een kwalitatieve en kwantitatieve inhoudsspecificatie, voor een algemene huisartsgeneeskundige kennistoets geformuleerd.

Het eerste deel van de inhoudsvaliditeitsstudie is er expliciet op gericht de keuzes en de procedures te verantwoorden die bij de blauwdrukontwikkeling zijn gemaakt, zodat kan worden nagegaan in hoeverre deze voldoet aan de inhoudelijke eisen. De blauwdrukontwikkeling en het resultaat ervan worden beschreven in paragraaf 2.3.

Het tweede deel van het onderzoek naar de inhoudsvaliditeit betreft een *empirische toetsing van de ontwikkelde blauwdruk*. In de blauwdruk is het kennisdomein onder meer vastgelegd in een bepaalde verdeling van vragen over onderwerpen. Daarmee geeft de blauwdruk een inhoudelijke stratificatie van het kennisdomein. Het was niet bekend in hoeverre deze stratificatie van invloed was op de toetsbetrouwbaarheid. Verschillende onderwerpen zouden

echter een verschillend effect op de betrouwbaarheid van de toets kunnen hebben (Jarjoura & Brennan, 1982). Dit zou aanleiding kunnen geven tot wijziging van de blauwdruk. Het onderzoek wordt gerapporteerd in de studie naar de inhoudsvaliditeit van de toets, omdat de resultaten ervan werden beschouwd als onderbouwing van de inhoudsvaliditeit van de kennistoets. In paragraaf 2.4 wordt meer in detail op dit onderdeel van de valideringsstudie ingegaan.

2.2.2. Bepaling constructvaliditeit

Constructvaliditeit betreft de mate waarin empirische bevindingen consistent zijn met theoretisch afgeleide hypothesen over de concepten (of constructen) die worden gemeten (Cronbach & Meehl, 1955; Cronbach, 1988, 1989). In hoofdstuk 1 is op basis van recente cognitief-psychologische theorievorming over expertise-ontwikkeling (Schmidt et al, 1990), het kennisbegrip dat in de algemene kennistoets aan de orde wordt gesteld, omschreven als de kennis die de beginnende huisarts na voltooiing van de huisartsopleiding, tot zijn beschikking moet hebben om zelfstandig praktijk te kunnen voeren. De constructvalidering is gericht op het aantonen van verschillen in toetsscores tussen groepen huisartsen-in-opleiding en huisartsen met een verschillend opleidings- en ervaringsniveau. Daartoe is een algemene kennistoets voorgelegd aan meerdere groepen aanstaande huisartsen en praktiserende huisartsen die in opleidings- en ervaringsniveau van elkaar verschilden. In paragraaf 2.5 wordt de constructvalidering van de toets nader beschreven.

2.3. BLAUWDRUKONTWIKKELING

2.3.1. Methode blauwdrukontwikkeling

De gebruiksfuncties van de toets dienen richtinggevend te zijn voor de criteria waaraan de blauwdruk moest voldoen. In de literatuur wordt aanbevolen om deze criteria, en daarmee de indelingssystematiek voor de te ontwikkelen toets, door de toetsontwikkelaar, in samenspraak met de gebruiker, zelf te laten bepalen (Thorndike, 1982). De criteria moeten corresponderen met zogenaamde 'critical abilities' voor de beroepsuitoefening, zodat de inhoud van de toets maximaal aansluit op de dagelijkse praktijk (Kane, 1982). De standaardmethode voor het bepalen van deze 'critical abilities' is om ervaren beroepsbeoefenaren te laten besluiten welke 'abilities' het meest wezenlijk zijn en geëvalueerd moeten worden (Kane, 1986). Deze procedure is bij de ontwikkeling van de blauwdruk gevolgd.

Op geleide van de eerder geformuleerde gebruiksfuncties voor de kennistoets werden de volgende criteria bepaald waaraan de blauwdruk zou moeten voldoen (Pollemans et al, 1993). Om de screenings- en selectiefunctie van de kennistoets te kunnen waarmaken moet de blauwdruk:

- het totale huisartsgeneeskundig kennisdomein dekken;
- ertoe leiden dat de kennistoetsen een medisch-inhoudelijk voldoende breed spectrum omvatten;
- bij voorkeur aansluiten bij andere indelingen (bijvoorbeeld bij indelingen die worden gebruikt in andere toetsen of bij morbiditeitsregistraties).

Om de kennistoets zinvol te kunnen toepassen als *educatief hulpmiddel*, moet de blauwdruk:

- méér dan alleen het klinisch handelen en zeker meer dan alleen het somatisch klinisch handelen omvatten;
- herkenbaar zijn voor de huisarts, dat wil zeggen uit onderwerpen moeten bestaan die een voorstelbare inhoud hebben voor de huisarts;
- zoveel mogelijk aansluiten bij de wijze waarop in de dagelijkse praktijk kennis wordt verworven.

Daarnaast werd nog een *praktisch criterium* geformuleerd, namelijk dat de blauwdruk:

- uit categorieën moet bestaan, waarover een voldoende aantal toetsvragen geformuleerd kan worden.

Elders gehanteerde indelingen sluiten inhoudelijk meestal direct aan bij het gegeven onderwijs, of volgen indelingen naar lichaamstracti of medische specialisaties (Verwijnen et al, 1982; Pisacano et al, 1989; Lockie, 1990). De diversiteit van de geformuleerde criteria maakte echter duidelijk dat het huisartsgeneeskundig kennisdomein niet in een beperkt aantal elkaar uitsluitende categorieën is te vangen. Daarom werden in de blauwdruk verschillende indelingsprincipes naast elkaar toegepast in de vorm van een classificatie met een hoofdingeling en meerdere nevenindelingen, die met elkaar moesten waarborgen dat alle relevante perspectieven werden gedekt (van Leeuwen e.a., 1992; van Leeuwen et al, in voorbereiding). Op basis van de criteria is vervolgens een kwalitatieve en kwantitatieve specificatie van de inhoud van een algemene huisartsgeneeskundige kennistoets ontwikkeld: de Blauwdruk voor de Huisartsgeneeskundige Kennistoets (Pollemans e.a., 1993; Pollemans e.a., 1994).

Verondersteld werd dat de acceptabiliteit van een algemene kennistoets voor huisartsen nauw samenhangt met de mate waarin de inhoud en de structuur ervan worden onderschreven door de (potentiële) gebruikers. De voorstellen voor definities van blauwdrukcategoryen en subcategoryen, en de verdeling van aantallen vragen over de verschillende onderdelen, inclusief de beargumentering van de voorstellen, werden daarom in een stapsgewijze *consensusprocedure* met een panel van experts op het gebied van huisartsgeneeskundige scholing en toetsing vastgelegd. Dit panel bestond uit specifiek deskundigen op het gebied van huisartsopleiding en -nascholing.¹ Met het panel werd in een aantal ronden consensus bereikt over de verschillende indelingsprincipes in de kennistoets.

De keuzes die ten grondslag lagen aan de samenstelling van de blauwdruk, inclusief hun argumentatie, werden aan het panel schriftelijk toegelicht (zie Bijlage 1 voor de vragenlijst betreffende inhoud en structuur van de huisartsgeneeskundige kennistoets). Vervolgens werden discussiebijeenkomsten georganiseerd met elk van de drie groeperingen die in het panel zitting hadden. Suggesties die tijdens de bijeenkomsten werden gedaan voor aanpassingen, aanvullingen of veranderingen werden in de definitieve blauwdruk verwerkt.

¹ In de consensusdiscussie waren betrokken:

- de Commissie Deskundigheidsbevordering van het NHG, met vertegenwoordigers van het beroepsveld die speciaal zijn belast met de voorbereiding van het landelijke deskundigheidsbevorderingsbeleid. In de Commissie participeerden vertegenwoordigers van NHG, LHV, O&O, WOK en huisartsopleiding;
- de acht hoofden van de huisartsopleiding: de beleidsverantwoordelijken voor voor de landelijke toetsing binnen de huisartsopleiding; en
- de acht afgevaardigden van de instituten in het Samenwerkingsverband die tot 1993 inhoudelijk verantwoorde-lijk waren voor de uitvoering van de landelijke toetsing en evaluatie in de huisartsopleiding.

2.3.2. Resultaat blauwdrukontwikkeling

- hoofdingeling

Bij de keuze van een indelingsprincipe voor de hoofdingeling van de blauwdruk speelde een aantal overwegingen een rol. De hoofdingeling moest terugkeren in de hoofdstukindeling van de kennistoets om te zorgen voor een optimale huisartsgeneeskundige 'herkenbaarheid' van de toets. Een van de consequenties daarvan was dat de hoofdingeling een duidelijk vakinhoudelijk perspectief moest weerspiegelen, dat tevens als basis voor zinvolle feedback over toetsresultaten kon worden gebruikt. Dat uitgangspunt hield weer in dat alle potentiële toetsvragen in elk geval in één van de categorieën van de hoofdingeling, moesten kunnen worden ondergebracht.

Het enige indelingsprincipe dat specifiek op de huisartsgeneeskunde is gericht, is de International Classification for Primary Care (ICPC) (Lamberts & Wood, 1987). De ICPC kent 17 elkaar onderling niet overlappende huisartsgeneeskundig-inhoudelijke hoofdstukken, die elk zijn onderverdeeld naar klachten en aandoeningen. Het systeem is speciaal ontwikkeld voor huisartsgeneeskundige morbiditeitsregistratie (Schadé & Lamberts, 1989).

Met de keuze van de indeling van klachten en aandoeningen volgens de ICPC als uitgangspunt voor de hoofdingeling van de blauwdruk, wordt tegemoetgekomen aan een aantal criteria: de ICPC is een indelingsprincipe dat herkenbaar is voor de beroepsgroep, het laat toe verbanden te leggen tussen morbiditeit, onderwijs en toetsgegevens, en de ICPC omvat in principe het huisartsgeneeskundige domein voor zover het klachten en aandoeningen betreft. Met behulp van de ICPC kan klacht- en diagnosegebonden kennis goed worden ingedeeld. Het systeem voorziet echter niet in een huisartsgeneeskundig-theoretische component, waarmee kennis kan worden gerubriceerd die niet direct klacht- of aandoeninggebonden is. Voorbeelden hiervan zijn de kennis over methoden van wetenschappelijk onderzoek, epidemiologie, concepten zoals gezinsgeneeskunde, somatische fixatie, theorievorming rondom patientenvoorlichting, of wet- en regelgeving op voor de huisarts relevante terreinen.

Als hoofdingelingsprincipe voor de blauwdruk is daarom gekozen voor de medisch-inhoudelijke hoofdstukken van de ICPC, aangevuld met een huisartsgeneeskundig-theoretische component.

Hoofdingeling blauwdruk

Medisch-inhoudelijke hoofdstukken ICPC

Huisartsgeneeskundig-theoretisch hoofdstuk

De bepaling van de getalsmatige verdeling van vragen over de blauwdrukcategorieën gebeurde in overleg met het expertpanel. De volgende factoren werden daarbij betrokken:

- gegevens uit morbiditeitsonderzoeken,
- variatie in klachten en aandoeningen binnen een bepaalde categorie,
- het belang van parate kennis bij de huisarts over het betreffende onderwerp,
- schatting van de mogelijkheden om voldoende verschillende kennistoetsvragen te kunnen samenstellen.

Gekozen werd voor een maximale toetsomvang van 160 items. Dat betekent een gemiddelde toetstijd van ongeveer twee uur. Hiervoor werd gekozen op grond van de ervaringen met de acceptabiliteit van deze toetstijd in de huisartsopleiding (Pollemans e.a., 1991) en om praktische redenen (de beschikbaarheid van voldoende kwalitatief goede toetsvragen). Gekozen werd verder voor vaste aantallen vragen per categorie. Ook deze keuze werd vooral

door praktische overwegingen ingegeven: de hoofdingeling van de blauwdruk zou terugkeren als hoofdstukindeling in de toets, en een vaste verdeling van vragen per hoofdstuk leek de herkenbaarheid van de toets te bevorderen.

In tabel 2.1 wordt de hoofdingeling van de blauwdruk weergegeven met het gewicht dat alle categorieën zouden moeten hebben, uitgedrukt in aantal vragen per categorie.

Tabel 2.1. Hoofdingeling blauwdruk.

Categorieën	Aantal items	Percentage
Algemeen en niet gespecificeerd (ICPC A)	8	5
Bloed en bloedvormende organen (ICPC B)	2	1
Tractus digestivus (ICPC D)	10	6
Oog (ICPC F)	6	4
Oor (ICPC H)	6	4
Tractus circulatorius (ICPC K)	16	10
Bewegingsapparaat (ICPC L)	18	11
Neurologie (ICPC N)	8	5
Psychische problemen (ICPC P)	8	5
Tractus respiratorius (ICPC R)	16	10
Huid en subcutis (ICPC S)	12	7
Endocriene klieren, metabolisme, voeding (ICPC T)	6	4
Urologie (ICPC U)	6	4
Zwangerschap, bevallingen, anticonceptie (ICPC W)	6	4
Vrouwelijke geslachtsorganen, incl. borsten (ICPC X)	6	4
Mannelijke geslachtsorganen, incl. borsten (ICPC Y)	6	4
Huisartsgeneeskundig-theoretische onderwerpen	20	12
Totaal	160	100

Uit tabel 2.1 blijkt dat alle ICPC-hoofdstukken als categorie terug te vinden zijn in de hoofdingeling van de blauwdruk, behalve ICPC-hoofdstuk Z (sociale problemen). In de praktijk blijken de kennisaspecten die onder deze categorie vallen, te herleiden tot kennis over (sociaal geïnduceerde) individuele problematiek en tot kennis over maatschappelijke problematiek. De eerste aspecten zijn niet goed af te scheiden van de kennis die onder hoofdstuk P (psychische problematiek) aan de orde komt. Kennis over maatschappelijke problematiek is ook onder de huisartsgeneeskundig-theoretische blauwdrukcategory te vatten. Daarom is het ICPC-hoofdstuk Z niet als aparte categorie in de blauwdruk opgenomen, maar vallen kennistoetsvragen over sociale problematiek in de categorie die is gebaseerd op ICPC-hoofdstuk P (voor zover het kennis betreft over individuele problematiek), ofwel in de categorie over huisartsgeneeskundig-theoretische onderwerpen (voor zover het kennis over niet-persoonsgebonden, sociale problemen betreft).

Bij het samenstellen van kennistoetsen op geleide van de blauwdruk worden de toetsvragen zo veel mogelijk gespreid over de inhoud van het betreffende ICPC-hoofdstuk. Dit om te voorkomen dat de hoofdstukken in de toets eenzijdig met vragen over bepaalde ziektebeelden of klachten worden gevuld.

Voor de huisartsgeneeskundig-theoretische onderwerpen zijn richtgetallen voor de opname van aantallen vragen in de kennistoets afgesproken. Deze onderverdeling is in tabel 2.2 weergegeven.

Tabel 2.2. *Onderverdeling huisartsgeneeskundig-theoretische onderwerpen in blauwdruk.*

Onderwerp	Aantal vragen
<i>Concepten van hulpverlening</i> preventie: somatische fixatie aandoeningen gezinsgeneeskunde theorie van consultvoering	2-6
<i>Wetenschap</i> epidemiologie besliskunde methodologie/statistiek wetenschapsfilosofie	4-8
<i>Maatschappelijke aspecten</i> sociale problemen regelgeving: tuchtrecht gezondheidsrecht sociale wetgeving medische ethiek organisatie en financiering: praktijk beroepsgroep gezondheidszorg	8-12
Totaal	20

Tabel 2.2 laat zien dat voor de verschillende aspecten in de theoretische categorie in de blauwdruk geen gespecificeerde onderverdeling naar aantallen vragen is gemaakt. Ook per categorie is de verdeling niet exact, maar met zekere onder- en bovenmarges aangegeven. Hiervoor is gekozen om een voldoende flexibiliteit bij de samenstelling van toetsen te kunnen garanderen. Evenals bij de andere, op ICPC-hoofdstukken gebaseerde, blauwdrukcategorieën wordt bij de samenstelling van toetsen rekening gehouden met een evenwichtige verdeling van vragen over de onderdelen van de verschillende subcategorieën.

- *nevenindelingen*

Omdat met het gebruik van de ICPC nog niet aan alle blauwdrukcriteria kon worden voldaan, werd een aantal nevenindelingen vastgesteld. Dit zijn classificaties die moeten garanderen dat de kennistoets ook vanuit andere perspectieven dan het hoofdinidelings-perspectief, een adequate dekking biedt van het huisartsgeneeskundig kennisdomein. Met behulp van het opnemen van meerdere nevenindelingen kan ook tegemoet worden gekomen aan onderwerpen die in de toekomst van toenemend belang voor de huisarts worden. De nevenindelingen zijn vastgesteld in overleg met het expertpanel. Besloten werd tot het opnemen van de volgende vier nevenindelingen.

Nevenindelingen blauwdruk
 Leeftijdscategorieën patiënten
 Aspecten van het consult
 Urgentie van het probleem (spoedgevallen)
 Beloop van de aandoening (chronische aandoeningen)

Belangrijk criterium bij de keuze voor nevenindelingen was dat daarmee de garantie gegeven kon worden dat wezenlijke aspecten van de huisartsgeneeskundige zorg, die speciale kennis en kunde van de huisarts vergen, in de kennistoets zijn vertegenwoordigd. De voorkeur voor een nevenindeling naar chronische ziekten werd ook ingegeven door het gewenste accent op de zorg van de huisarts voor chronisch zieken.

Voor de nevenindelingen werden streefgetallen opgesteld voor maximaal of minimaal in de kennistoets op te nemen aantallen vragen. Ook deze getallen waren gebaseerd op een combinatie van factoren als gegevens uit morbiditeitsstatistieken, variatie in klachten en aandoeningen, het belang van parate kennis bij de huisarts, geschatte moeilijkheidsgraad en opmerkingen en wensen van het expertpanel. Wat betreft de nevenindelingen verkreeg het aangeven van streefpercentages de voorkeur boven het aangeven van vaste aantallen vragen per nevenindeling. Ook hiervoor golden praktische argumenten, die vooral te maken hadden met een noodzakelijk mate van flexibiliteit die bij het construeren van toetsen zou moeten blijven gelden.

Tabel 2.3. *Nevenindelingen blauwdruk.*

Nevenindeling	Percentage vragen/casus
<i>Leeftijdscategorieën patiënten</i> < 15 jaar tussen 15 en 75 jaar > 75 jaar	10-25% van het aantal casus 50-80% van het aantal casus 10-25% van het aantal casus
<i>Aspecten van het consult</i> diagnostiek beleid medicamenteus niet-medicamenteus epidemiologie rest	minstens 40% van het aantal vragen minstens 40% van het aantal vragen minstens 20% minstens 20% maximaal 10% van het aantal vragen maximaal 10% van het aantal vragen
<i>Spoedgevallen</i>	5-10% van het aantal casus
<i>Chronische aandoeningen</i> checklist	10-25% van het aantal casus

Uit tabel 2.3 blijkt dat sommige nevenindelingen betrekking hebben op het aantal casus, en andere op het aantal vragen in de toets. De leeftijdscategorieën zijn gebaseerd op de indeling van Van de Lisdonk (Van de Lisdonk, 1991). De leeftijdscategorie < 15 jaar werd ingevoerd om ervoor te zorgen dat de kennistoetsen voldoende vragen bevatten die typische kinderpathologie representeren; de categorie > 75 jaar om voldoende vragen in te toetsen te krijgen die typische ouderdomspathologie representeren.

Voor de subcategorieën onder de nevenindelingen naar 'aspecten van het consult' en 'spoedgevallen' werden definities opgesteld die op hun bruikbaarheid als indelingscriterium voor toetsvragen werden getoetst in een *inter-indelaars-betrouwbaarheidsonderzoek*. Onderzocht werd in hoeverre een viertal huisartsen overeenstemde in het indelen van toetsvragen volgens deze definities. Deze huisartsen hadden allen ervaring met het formuleren van kennistoetsvragen op het gebied van de huisartsgeneeskunde. Hen werd gevraagd een selectie van casus (c.q. items) van twee landelijke kennistoetsen (totaal aantal casus 164) in te delen volgens de voorgestelde definities. Hen werd ook gevraagd de opgestelde definities en omschrijvingen te becommentariëren. De casus werden aan ieder van de vier huisartsen in een andere volgorde aangeboden. De inter-indelaars-betrouwbaarheid werd op twee manieren berekend en weergegeven als kappa (Hubert) en kappa (Ross). Voor de nevenindeling 'aspecten van het consult' bleek de kappa (Hubert) 0.79 en de kappa (Ross) 0.73. Dat betekende dat er sprake was van een redelijke tot goede overeenstemming tussen de vier huisartsen. Voor de nevenindeling 'spoedgeval' bleek de kappa (Hubert) 0.74 en de kappa (Ross) 0.76. Ook hier was de conclusie dat er sprake was van een redelijke tot goede overeenstemming (Düsmen, 1993). Op basis van de commentaren van de vier huisartsen werden de definities nog enigszins aangescherpt.

Voor de nevenindeling naar chronische aandoeningen kon geen sluitende definitie worden gevonden om de casus te kunnen indelen. Daarom werd voor deze nevenindeling een lijst van aandoeningen opgesteld, die doorgaans een chronisch karakter hebben. Deze lijst is een uitbreiding van de lijst volgens Lamberts (Lamberts, 1991). In deze lijst wordt een aantal ziekten als chronisch aangemerkt, waarbij het criterium vooral de beperking van de ADL is. De bestaande lijst is aangevuld met een aantal chronische ziekten die naar het oordeel van de geraadpleegde experts (hoofden en stafleden huisartsinstituten, leden Werkgroep Kennis Samenwerkingsverband) ook voor opname als chronische aandoening in de huisartsgeneeskundige kennistoets in aanmerking kwamen. Criterium was uiteindelijk dat het aandoeningen betroffen die continue zorg van de huisarts behoeven. De definities en de checklist zijn in Bijlage 2 opgenomen.

2.4. EMPIRISCHE INHOUDSVALIDERING BLAUWDruk

De blauwdruk specificeert de toets in 17 inhoudsgebieden. Het was onbekend of, en zo ja in welke mate deze inhoudsgebieden elk op zich een verschillende bijdrage aan de betrouwbaarheid van de totale toets leverden. Denkbaar was dat de inhoudsgebieden in dit opzicht van elkaar zouden verschillen. Als de inhoudsgebieden elk voor zich een zeer verschillende bijdrage aan de totale toetsbetrouwbaarheid zouden leveren, zou dit aanleiding kunnen geven tot wijzigingen van de relatieve omvang van de verschillende inhoudsgebieden in de blauwdruk en daarmee consequenties hebben voor de inhoudsvaliditeit van de toets. De vraag was dus of een andere verdeling van de toetsvragen over de onderwerpen een hogere toetsbetrouwbaarheid zou opleveren.

2.4.1. Methode: multivariate generaliseerbaarheidsanalyse

Het design van de beschikbare gegevens (scores van deelnemers op toetsen die waren samengesteld volgens de blauwdruk met vaste inhoudsgebieden) maakte het mogelijk de

bijdrage van de inhoudsgebieden aan de totale toetsbetrouwbaarheid te onderzoeken met behulp van een multivariate generaliseerbaarheidsanalyse (Brennan, 1983). Hiermee kon voor elk inhoudsgebied worden bepaald bij welk aantal items de betrouwbaarheid van de toets maximaal wordt (bij de gegeven totale toetslengte). Vervolgens wordt nagegaan in welke mate de betrouwbaarheid van de toets verandert bij een andere verdeling van itemaantallen over de onderwerpen. (Voor een toelichting op de begrippen die hierna worden gebruikt, wordt verwezen naar Bijlage 3 waarin wordt ingegaan op de generaliseerbaarheidstheorie die het kader voor de analyses vormde; voor de formules zie Bijlage 4.)

Conform de uitgangspunten van de (multivariate) generaliseerbaarheidstheorie (Cronbach et al, 1972; Brennan, 1983) werd een G-studie uitgevoerd per blauwdruk-onderwerp. De volgende variantiebronnen werden per onderwerp onderscheiden.

Variantiebronnen

P (personen)

I (items)

PI (interactie tussen personen en items)

De analyses zijn uitgevoerd op het empirisch materiaal, dat beschikbaar was na de afname van drie volgens de blauwdruk samengestelde toetsen bij gevestigde huisartsen. Voor het samenvoegen van gegevens over drie toetsen is gekozen om tot stabielere uitspraken te kunnen komen (Brennan, 1983). Uitgangspunt voor de berekeningen vormden de gegevens van de huisartsen die deelnamen aan de kennistoetsen van juni 1991, oktober 1991 en februari 1992.²

Per toets werden met het programma GENOVA (Brennan, 1983) per onderwerp variantiecomponenten berekend voor de genoemde variantiebronnen, en vervolgens gemiddeld over de drie toetsen. De covarianties van de onderwerpen werden eveneens per toets berekend en vervolgens gemiddeld voor alle deelnemende huisartsen (in totaal 523). De multivariate generaliseerbaarheidsanalyse werd uitgevoerd met het programma MGENOVA (van der Vleuten, 1991) op basis van de gemiddelde variantiecomponenten en de gemiddelde covarianties van de onderwerpen. Vervolgens konden D-studies worden uitgevoerd ten behoeve van de uiteindelijke betrouwbaarheidsschatting.

2.4.2. Resultaat multivariate generaliseerbaarheidsanalyse

Het resultaat van de analyses is opgenomen in tabel 2.4 waarin, gegeven de beschikbare data, een optimale verdeling van items over de onderwerpen is weergegeven. De tabel representeert het resultaat van de D-studies uitgevoerd op basis van de goed-min-foutscores van de huisartsen die deelnamen aan drie op de blauwdruk gebaseerde kennistoetsen. In totaal gaat het om analyses berekend over toetsscores van 523 huisartsen; de toetslengte bedroeg steeds 160 items (Düsmen, 1992).

² Deze drie toetsen waren de eerste kennistoetsen die waren samengesteld op geleide van de blauwdruk. Aan alle toetsen namen naast huisartsen-in-opleiding ook ervaren huisartsen deel. De toetsen van juni en oktober 1991 moeten worden beschouwd als een soort 'voorlopers' of pilotstudies van de definitieve toets van februari 1992. De analyses met betrekking tot de constructvalidering en de betrouwbaarheid van de kennistoets voor ervaren huisartsen, zijn alle gebaseerd op deze toets.

De blauwdruk-onderwerpen en de feitelijke aantallen vragen per onderwerp, staan vermeld in de eerste twee kolommen van de tabel. In kolom 3 is het resultaat van de G-studie weergegeven: de absolute bijdrage per onderwerp aan de bedoelde variantie. Deze is in de vierde kolom als relatieve bijdrage aan de variantie uitgedrukt (variantie per onderwerp gedeeld door de totale variantie).

In beide laatste kolommen is het resultaat van de D-studies opgenomen. In kolom 5 is aangegeven wat het optimale aantal items per onderwerp zou zijn, als alleen wordt gekeken naar de stabiliteit van de onderlinge verschillen tussen de deelnemers (het zogenaamde *normgeoriënteerde of relatieve perspectief*). In kolom 6 tenslotte is per onderwerp het optimale aantal items opgenomen als ook de beheersing van de stof een rol speelt (het zogenaamde *domeingeoriënteerde of absolute perspectief*) (Norcini & Swanson, 1989; Suen, 1990).

Tabel 2.4. Resultaat multivariate generaliseerbaarheidsanalyse algemene kennistoets voor huisartsen.

Blauwdrukcategorie	Aantal items	Absolute bijdrage aan bedoelde variantie	Relatieve bijdrage aan bedoelde variantie	Optimaal aantal items (norm-georiënteerd)	Optimaal aantal items (domein-georiënteerd)
Algemeen en niet gespecificeerd	8	0.0005	0.0528	8.14	7.71
Bloed en bloedvormende organen	2	0.0001	0.0144	1.95	1.95
Tractus digestivus	10	0.0005	0.0577	9.4	9.65
Oog	6	0.0003	0.0284	6.35	6.36
Oor	6	0.0002	0.0243	5.84	5.64
Tractus circulatorius	16	0.001	0.1148	15.77	14.99
Bewegingsapparaat	18	0.001	0.11	19.04	18.73
Neurologie	8	0.0006	0.069	8.87	9.03
Psychische problemen	8	0.0004	0.0478	8.58	8.74
Tractus respiratorius	16	0.0009	0.0988	15.35	15.47
Huid en subcutis	12	0.0006	0.07	11.04	10.82
Endocr. klieren/metabolisme/voeding	6	0.0003	0.0378	5.47	5.77
Urologie	6	0.0007	0.076	5.74	6.62
Zw.schap/bevallingen/anticonceptie	6	0.0002	0.0259	6.39	6.68
Vrouw. geslachtsorg. incl. borsten	6	0.0004	0.0476	6.48	6.48
Mann. geslachtsorg. incl. borsten	6	0.0003	0.0366	6.4	6.23
Theoretische onderwerpen	20	0.0008	0.0882	19.21	19.12
Totaal	160	0.009	1.00	160	160

Tabel 2.4 laat zien dat slechts geringe verschillen tussen de feitelijk gehanteerde en de vanuit betrouwbaarheidsoogpunt optimale aantallen items per blauwdrukcategorie werden gevonden. Dit wijst erop dat de 17 blauwdrukonderwerpen nauwelijks van elkaar verschillen wat betreft hun bijdrage aan de betrouwbaarheid aan de toets. Het resultaat van een D-studie waarin de aantallen items per onderwerp zijn 'geoptimaliseerd' conform de bevindingen uit de multivariate generaliseerbaarheidsstudie, wordt weergegeven in tabel 2.5.

Tabel 2.5. Betrouwbaarheidscoëfficiënten algemene kennistoets bij verandering van aantallen items per onderwerp.

Betrouwbaarheidsgegevens	Toetssamenstelling volgens blauwdruk	Optimale toetssamenstelling (resultaat D-studie)
Totale toetsvariantie	.00907	.00927
Relatieve errorvariantie	.00355	.00357
Absolute errorvariantie	.00441	.00445
Normgeoriënteerde betrouwbaarheidscoëfficiënt	.72	.72
Domeingeoriënteerde betrouwbaarheidscoëfficiënt	.67	.67

Een andere verdeling van items over de blauwdrukonderwerpen, leidt niet tot andere betrouwbaarheidsindices. Dit geldt zowel voor het normgeoriënteerde (relatieve) perspectief, als voor het domeingeoriënteerde (absolute) perspectief op de betrouwbaarheid van toetsscores.

Voor de toets betekent het dat de verdeling van items volgens de blauwdruk, die is gebaseerd op inhoudelijke argumenten, vanuit betrouwbaarheidsoogpunt niet voor verbetering vatbaar is. Wijziging van de blauwdruk zou de betrouwbaarheid niet of nauwelijks kunnen verbeteren. Relatieve verlenging of verkorting van de onderwerpen in de toets maakt, met andere woorden, de totaalscore niet betrouwbaarder.

2.5. CONSTRUCTVALIDERING KENNISTOETS

Het kennisbegrip dat in de huisartsgeneeskundige kennistoets aan bod komt, is in hoofdstuk 1 omschreven als praktisch relevante kennis die aan het eind van de huisartsopleiding moet worden beheerst. Het vormt als het ware de gemeenschappelijke basis waarover beginnende beroepsbeoefenaren moeten beschikken. Deze basis wordt verworven met de ervaring tijdens hun opleiding in de huisartsgeneeskunde. De expertise die met het toenemen van de ervaring als praktiserend huisarts wordt ontwikkeld, is gebaseerd op de integratie van individuele, via patiëntencontacten verworven ervaringskennis met algemene, vooral door opleiding en scholing verworven conceptuele kennis (Schmidt et al, 1990; Boshuizen & Schmidt, 1993). Het kennisbegrip in kennistoetsen die zich richten op ervaren huisartsen, omvat beide aspecten.

Medische studenten vormen een groep aanstaande artsen, die niet is gedifferentieerd naar (huisarts)geneeskundige voorkeur. Als de kennistoets inderdaad typisch huisartsgeneeskundige kennis meet, mag worden verwacht dat het gemiddelde niveau van de (bijna-)basisartsen lager is dan dat van pas beginnende huisartsen-in-opleiding.

De huisartsopleiding beoogt op te leiden tot het niveau dat nodig is om het LHV-Basistakenpakket (Springer, 1983) in de praktijk uit te kunnen voeren (Pollemans & Tan, 1990). Als de kennistoets inderdaad in staat is om deze kennis te meten, moet dit, naarmate men meer met de opleiding is gevorderd, in gemiddeld hogere scores tot uiting komen (van Leeuwen, in voorbereiding). Er werd van uitgegaan dat de ontwikkeling van kennis over theoretische

concepten en recente inzichten en nieuwe ontwikkelingen op het vakgebied het meest geconcentreerd plaatsvindt tijdens de huisartsopleiding. Daarom was de verwachting dat het algemene kennisniveau aan het eind van de opleiding het hoogst zou zijn. Daarbij werd er op theoretische gronden van uitgegaan dat ervaren huisartsen een kennisbegrip hebben dat zich steeds meer kenmerkt door 'kennis over patiënten' dan door 'kennis over (theoretische) problemen' (van Leeuwen et al, in voorbereiding).

Individueel verworven patiëntgebonden ervaringskennis is per definitie afhankelijk van de verschillende praktijkervaringen van huisartsen (Boshuizen, 1989). Het leek daarom onwaarschijnlijk dat de kennistoets in staat zou zijn ook deze kennis, in de totaalscores op groepsniveau, zichtbaar te maken. De verwachting leek eerder gerechtvaardigd dat gevestigde huisartsen gemiddeld minder hoge scores zouden bereiken naarmate hun opleiding verder achter hen lag (Grant & Marsden, 1988). Daarbij werd verwacht dat de individueel verschillende ervaringen elkaar, op het niveau van de totale toets, als het ware zouden 'uitmiddelen'.

2.5.1. Methode

Om aanwijzingen te verkrijgen over de constructvaliditeit van de algemene huisartsgeneeskundige kennistoets, werd gebruik gemaakt van de gegevens die in februari 1992 werden verzameld bij de afname van een volgens de blauwdruk samengestelde toets van 160 vragen bij groepen ervaren huisartsen, huisartsen-in-opleiding in diverse opleidingsfasen en groepen co-assistenten huisartsgeneeskunde (zie Bijlage 5 voor een voorbeeldpagina uit de algemene huisartsgeneeskundige kennistoets van februari 1992). De toetsafname vond plaats in dezelfde tijdperiode als een reguliere toetsafname in de huisartsopleiding. De gegevens werden verwerkt en geanalyseerd door het Uitvoerend Bureau van het landelijk Samenwerkingsverband van de Universitaire Huisartsopleidingen volgens gangbare procedures. Ten behoeve van de itemanalyse werden per toetsvraag eerst percentages goede en foute antwoorden, plus percentage vraagtekens berekend, alsmede itemtestcorrelaties en discriminatie-indices. Deze gegevens maken duidelijk in hoeverre de individuele vragen in de toets zich 'afwijkend' gedragen van de toets als geheel, en in hoeverre individuele vragen consistent goed beantwoord worden door deelnemers die ook de rest van de toets goed beantwoorden, c.q. verkeerd worden beantwoord door deelnemers die ook op de rest van de toets geen goed resultaat behalen (Hamaker, 1980; van Hessen & Imbos, 1987). In de itemanalyse werden deze psychometrische gegevens betrokken alsmede het commentaar van de deelnemers op de toetsvragen. Op grond hiervan zijn na afloop van de toets 14 van de 160 vragen vervallen. Alle nadere analyses zijn gebaseerd op de resterende kennistoets van 146 vragen. Conform de gebruikelijke procedure bij het Samenwerkingsverband is de gemiddelde goed-min-foutscore gehanteerd als toetsscore. Daarbij gelden correcte antwoorden voor 1 punt, wordt voor verkeerde antwoorden 1 punt afgetrokken, en tellen vraagtekens niet mee. De toetsscore is een percentage van de maximaal te bereiken score en wordt uitgedrukt in een getal dat ligt tussen -100 en +100.

2.5.2. Resultaat: scoreverloop en expertiseontwikkeling

De resultaten van de constructvalideringsstudie zijn opgenomen in tabel 2.6. De resultaten worden weergegeven voor de totale groepen co-assistenten, huisartsen-in-opleiding en huisartsen. De resultaten van de huisartsen-in-opleiding worden bovendien weergegeven voor

de eerstejaars- en de tweedejaars huisartsen-in-opleiding. De ervaren huisartsen worden in drie groepen verdeeld: een groep met minder dan tien jaar ervaring, een groep met tien tot twintig jaar ervaring en een groep met meer dan twintig jaar ervaring.

Tabel 2.6. Gemiddelde goed-min-foutcores van co-assistenten huisartsgeneeskunde, huisartsen-in-opleiding en ervaren huisartsen op de algemene kennistoets, afgenomen in februari 1992.

Algemeen kennisniveau	Gemiddelde goed-min-foutscore (+ std) (%)							
	co's	huisartsen-in-opleiding			ervaren huisartsen			
		1e jaar	2e jaar	totaal	< 10 jaar	10-20 jaar	> 20 jaar	totaal
	N 108	N 234	N 211	N 445	N 90	N 208	N 53	N 351
Totale toets (146 vragen)	23 (9)	37 (11)	45 (11)	41 (11)	48 (11)	44 (11)	41 (12)	45 (11)

Van de drie groepen die aan de toets deelnamen (co-assistenten, huisartsen-in-opleiding en huisartsen) is de gemiddelde goed-min-foutscore van de totale groep ervaren huisartsen het hoogste. Co-assistenten scoren gemiddeld aanzienlijk lager dan huisartsen-in-opleiding. De hoogste gemiddelde score wordt aangetroffen bij de groep huisartsen die het meest recent is afgestudeerd. Met toenemende ervaring nemen de gemiddelde scores van de ervaren huisartsen gestaag af. De verschillen tussen de onderscheiden opleidings- en ervaringsgroepen zijn statistisch significant ($p < 0.05$).

2.6. CONCLUSIES VALIDITEITSSTUDIE ALGEMENE HUISARTSGENEESKUNDIGE KENNISTOETS

De vraag die in dit hoofdstuk aan de orde is gesteld betreft de validiteit van de algemene huisartsgeneeskundige kennistoets. De inhoudsvaliditeit en de constructvaliditeit staan centraal. De algemene conclusie van de valideringsstudie is dat samenstelling van kennistoetsen op geleide van de ontwikkelde blauwdruk, in combinatie met de gehanteerde casuïstische vraagvorm, voldoende garantie vormt voor het een valide en praktijkrelevante toets. Hiervoor geldt de volgende argumentatie.

De blauwdruk voldoet aan de criteria die werden geformuleerd voor een verantwoorde inhoudsstructuur voor de kennistoets. Door de keuze voor een hoofdingeling en een viertal nevenindelingen dekt de blauwdruk het gehele huisartsgeneeskundige domein. Omdat de hoofdingeling de medisch-inhoudelijke onderwerpen van de ICPC volgt, wordt een voldoende breed spectrum omvat en wordt tevens aangesloten bij een gebruikelijke indelingssystematiek in de huisartsgeneeskunde. De blauwdruk omvat meer dan alleen somatisch klinisch handelen. Er is daartoe, naast de ICPC-categorieën, een apart onderdeel in de blauwdruk opgenomen, dat huisartsgeneeskundig-theoretische onderwerpen betreft. Hoofdingeling en nevenindelingen

representeren herkenbare thema's voor de huisarts. Omdat is gekozen voor een hoofdingeling, die is gebaseerd op een huisartsgeneeskundige morbiditeitsindeling waarin klachten zowel als aandoeningen als invalshoek fungeren, sluit de blauwdruk aan bij de wijze waarop in de dagelijkse praktijk kennis wordt verworven.

De representativiteit van de inhoudsstructuur van de blauwdruk voor het huisartsgeneeskundig kennisdomein, wordt ondersteund door experts op het gebied van het huisartsgeneeskundig onderwijs. Deze experts waren afkomstig uit de huisartsopleiding en uit de nascholing. Vanuit het oogpunt van de te bereiken betrouwbaarheid is de verdeling van items over de onderwerpen volgens de blauwdruk, optimaal.

Op grond van bovenstaande gegevens, wordt geconcludeerd dat de blauwdruk voldoende garantie oplevert om er inhoudsvalide toetsen op te baseren.

De algemene huisartsgeneeskundige kennistoets is in staat om verschillen in gemiddelde kennis aan te tonen tussen huisartsen in diverse fasen van hun opleiding of ervaring. De empirische bevindingen na afname van de toets bij, qua opleidings- en ervaringsniveau, verschillende groepen deelnemers, lieten significante verschillen zien tussen de scores van de deelnemersgroepen. De ervaren huisartsen bereikten daarbij de hoogste scores. Als meer gedetailleerd naar de scores van de verschillende groepen wordt gekeken, blijkt dat de hoogste scores worden gevonden bij de groepen die zich aan het eind van de opleiding, c.q. in het begin van hun beroeps carrière bevinden. De gemiddelde scores van huisartsen en huisartsen-in-opleiding in verschillende fasen van opleiding en ervaring, vertoonden een patroon dat op theoretische gronden kon worden verwacht. Deze bevindingen worden beschouwd als ondersteuning van de constructvaliditeit van de algemene huisartsgeneeskundige kennistoets.

De nadere specificatie van de kennistoets door de hoofdingeling en de nevenindelingen van de blauwdruk, staat ervoor garant dat toetsresultaten kunnen worden teruggerapporteerd volgens een praktijkrelevante systematiek. Inhoudelijk voldoet de toets daarom aan de eisen die voortvloeien uit het gewenste diagnostische gebruik van de toets als screeningsinstrument en als hulpmiddel om individuele sterke en zwakke kanten van het kennisniveau van ervaren huisartsen zichtbaar te maken.

Het uiteindelijke oordeel over de bruikbaarheid van de kennistoets voor deze functies, is mede afhankelijk van de gegevens die de studie naar de betrouwbaarheid van de toetsresultaten heeft opgeleverd. Hierop wordt in het volgende hoofdstuk ingegaan.

Hoofdstuk 3

BETROUWBAARHEID ALGEMENE HUISARTSGENEESKUNDIGE KENNISTOETS

SAMENVATTING

Toetsscores op zich zijn getallen zonder context, ze spreken niet voor zichzelf. Om er betekenis aan te kunnen toekennen, moeten ze worden geïnterpreteerd. Interpretatie is mogelijk als er informatie is over de betrouwbaarheid van de scores. De betrouwbaarheid kan vanuit drie perspectieven worden geïnterpreteerd: normgeoriënteerd (of relatief), domeingeoriënteerd (of absoluut) en beslissingsgeoriënteerd. Bij de interpretatie van scores vanuit een normgeoriënteerd perspectief geeft de score inzicht in de betrouwbaarheid van de rangorde van de scores, dat wil zeggen van de positie van de score ten opzichte van de scores van anderen. Vanuit het domeingeoriënteerde perspectief gaat het erom in hoeverre de score aangeeft hoe de toetsprestaties zich verhouden tot van te voren goed gedefinieerde taken, ofwel in welke mate het vakgebied wordt beheerst. Vanuit het beheersingsperspectief gaat het om de relatieve positie van de score ten opzichte van een cesuur of zak-slaaggrens.

Het gebruik van de generaliseerbaarheidstheorie in een betrouwbaarheidsstudie, maakt het mogelijk voor de drie verschillende perspectieven betrouwbaarheidsindices te schatten. Dat is van belang omdat de verschillende betrouwbaarheidsperspectieven samenhangen met bepaalde gebruiksfuncties van de toets. De betrouwbaarheidsstudies zijn daarom uitgevoerd op basis van deze theorie.

Betrouwbaarheidsschattingen werden gemaakt voor individuele scores, bij wisselende toetslengte, en voor groepsgegevens, bij variabele groepsgrootte en toetslengte.

Voor de individuele scores werden generaliseerbaarheids-, phi- en adjusted phi-coëfficiënten geschat plus standaardmeetfouten (SEM); voor de groepscores werden standaardmeetfouten vanuit het relatieve en het absolute perspectief als betrouwbaarheidsindicator gekozen.

De betrouwbaarheid van individuele scores, vanuit normgeoriënteerd (relatief) perspectief, van de algemene huisartsgeneeskundige kennistoets bedroeg .70 met een SEM van 6. Vanuit domeingeoriënteerd (absoluut) perspectief bedroeg de betrouwbaarheid .65 (SEM 7). Verdubbeling van de toets tot 320 items met een toetsduur van naar schatting vier uur zou een betrouwbaarheid opleveren van ruim .80 voor beide perspectieven, maar de SEM zou dan nog steeds 4, resp. 5 bedragen. Betrouwbare beslissingen (met een betrouwbaarheidscoëfficiënt van .80) kunnen worden genomen op basis van een cesuur die ligt onder een goed-min-foutscore van 35 of boven die van 55%. De betrouwbaarheid van gemiddelde groepsscores liet een veel gunstiger beeld zien. Vanuit normgeoriënteerd (relatief) perspectief is de SEM 1 vanaf een groepsgrootte van ongeveer 30 tot 50 deelnemers; vanuit domeingeoriënteerd (absoluut) perspectief is de SEM dan 3.

Geconcludeerd wordt dat voor screeningsdoelen, als het dus vooral om groepsgemiddelden gaat, de betrouwbaarheid van de algemene huisartsgeneeskundige kennistoets voldoende acceptabel is. Op individueel niveau moet rekening worden gehouden met een (te) grote foutenmarge om zonder meer zak-slaagconsequenties aan het toetsresultaat te kunnen verbinden. De betrouwbaarheid van beslissingen varieert met de plaats waar de cesuur wordt gelegd en is hoger naarmate deze verder van de gemiddelde score wordt gelegd. Omdat onduidelijk is welk minimumniveau aan kennis nog verenigbaar is met het goed functioneren als huisarts, is op grond van dit onderzoek geen goede uitspraak te doen over de meest optimale plaats van de cesuur voor voldoende of onvoldoende kennisbeheersing. Deze conclusie laat onverlet dat de toets voor educatieve doelen goed bruikbaar is, mits de marges waarin de 'ware' score kan liggen bij de toetsresultaten worden vermeld.

3.1. INLEIDING

Betrouwbaarheid van meetresultaten betreft de vraag naar de consistentie van meetresultaten bij herhaling van de toets. Deze is bepalend voor de gebruikswaarde van een meetinstrument. Personen die de ene keer hoog scoren zouden dat, onder vergelijkbare omstandigheden, een andere keer ook moeten doen. De betrouwbaarheid geeft dus aan in hoeverre variatie in scores te maken heeft met systematische verschillen tussen personen, of met andere zaken die tot variatie in de scores leiden (Swanson et al, 1987). In het algemeen geldt dat hoe consistent de resultaten over verschillende metingen zijn, des te betrouwbaarder de resultaten zijn. Betrouwbaarheid wordt daarom ook wel als een kwaliteitseis voor een meetinstrument beschouwd (van Berkel & Bax, 1990).

Het begrip betrouwbaarheid wordt in de literatuur beschreven vanuit een aantal meer of minder van elkaar verschillende uitgangspunten die elk leiden tot een bepaalde betrouwbaarheidstheorie. De belangrijkste zijn: de klassieke testtheorie en de daaraan verwante generaliseerbaarheidstheorie, en de itemresponstheorieën (Suen, 1990; Eggen & Sanders, 1993).

In de *klassieke testtheorie* wordt ervan uitgegaan dat een (waargenomen) score bestaat uit twee delen: de 'echte' of ware score, en een toevallig deel: de error of meetfout. De ware score is een hypothetische kwantiteit: het is de score die iemand zou behalen als hij een oneindig aantal keer op een bepaalde variabele werd gemeten. De meetfout omvat alle toevalsfactoren die het resultaat van een meting in de praktijk oncontroleerbaar beïnvloeden. Naarmate de meetfout groter is, is het meetresultaat minder betrouwbaar (Drenth & Sijsma, 1990). De klassieke testtheorie biedt een theoretisch model om de meetfout te kunnen schatten. Daartoe wordt een aantal aannames gedaan: dat er geen verband bestaat tussen het ware deel van de score en het deel dat door toevallige omstandigheden wordt bepaald; en dat de toevallige componenten die de score beïnvloeden, elkaar neutraliseren in het hypothetische geval van een oneindig aantal metingen. Daarnaast wordt in het theoretische model gebruik gemaakt van het feit dat niet één persoon een toets maakt maar dat meerdere personen dat doen, zodat gebruik kan worden gemaakt van de spreiding van de scores over de items (Eggen & Sanders, 1993). Onder deze assumpties kan worden aangetoond dat de totale variantie van een toets identiek is aan de som van de ware variantie en die van de toevalsvariantie (Drenth, 1975). Op basis van deze uitgangspunten wordt de betrouwbaarheid van een toets gedefinieerd als de ratio tussen de ware variantie en de totale variantie

(= waargenomen variantie, dat wil zeggen ware variantie plus foutenvariantie) van die toets. Dit levert dus een coëfficiënt van betrouwbaarheid op die waarden kan aannemen tussen 0 en 1. Als de betrouwbaarheid 0 is, is er geen ware variantie. Het instrument meet dan alleen meetfouten. Naarmate de variantie teweeggebracht door de meetfout geringer is, stijgt de betrouwbaarheid. Als de betrouwbaarheid 1 is, worden er geen meetfouten met het instrument gemeten. De ware en de totale variantie komen dan immers geheel met elkaar overeen (Eggen & Sanders, 1993).

Als de meetfout veel invloed op de waargenomen score heeft, dat wil zeggen bij een lage betrouwbaarheid, zal dat betekenen dat bij herhaling van de meting weinig kans bestaat dat dezelfde score wordt gevonden (Thomdike, 1971; Carmines & Zeller, 1985). De coëfficiënt waarmee de betrouwbaarheid wordt uitgedrukt is dan ook een maat voor de reproduceerbaarheid van de toetsscores. Idealiter wordt deze coëfficiënt bepaald door de variantie te bepalen van de scores op twee toetsafnames onder identieke omstandigheden. Dit is echter in de praktijk onmogelijk. In de klassieke testliteratuur worden daarom verschillende schattingsmethoden beschreven die elk leiden tot een bepaalde betrouwbaarheidscoëfficiënt (Meerling, 1981; Pollemans & Düsman, 1994). De betrouwbaarheid kan worden geschat met behulp van methoden die zijn gebaseerd op herhaalde toetsing, of met behulp van methoden waarin bewerkingen plaatsvinden op basis van de scores behaald op een enkele toets. Een voorbeeld van een methode waarbij slechts één meting wordt gebruikt, is de berekening van de betrouwbaarheid die is gebaseerd op itemanalyse. Dit leidt tot een zogenaamde coëfficiënt van itemconsistentie, ofwel homogeniteitscoëfficiënt: Cronbach's alpha (Drenth 1975).

De *generaliseerbaarheidstheorie* is eigenlijk een uitbreiding van de klassieke testtheorie en hanteert ook dezelfde uitgangspunten wat betreft de schatting van de betrouwbaarheid op basis van toetsscores (Cronbach et al, 1972; Brennan, 1983). De generaliseerbaarheidstheorie maakt het mogelijk zicht te krijgen op de invloed van verschillende bronnen die tot de meetfout bijdragen (Allal, 1988). Daarbij is het gebruiksdoel van de toets bepalend zowel voor de bron die wordt beschouwd als 'object van meting' (dit is de variantiebron die leidt tot de 'ware' of bedoelde variantie) als voor de bronnen die worden gerekend als bijdragend aan de meetfout (van der Vleuten & Wijnen, 1991). Dat betekent dat, afhankelijk van het gebruiksdoel van de toets, sommige bronnen wel en andere niet tot de meetfout worden gerekend. Afhankelijk van het gebruiksdoel kan dit leiden tot verschillende coëfficiënten van betrouwbaarheid (zie ook Bijlage 3).

In de *itemresponsstheorieën* wordt de betrouwbaarheid bepaald op basis van de score op afzonderlijke items (van der Linden, 1978). De betrouwbaarheid neemt af naarmate mensen met minder kennis een bepaald item goed beantwoorden, en mensen met meer kennis het item fout beantwoorden. Volgens deze theorieën zouden toetsen moeten bestaan uit items die op zichzelf, dus los van de toets-context, 'sterk' genoeg zijn om betrouwbare uitspraken op te baseren. Het gevolg hiervan zijn zogenaamde 'testonafhankelijke toetsscores' (Crocker & Algina, 1986). Dit maakt enerzijds vergelijking mogelijk tussen items, onafhankelijk van de toevallige groep personen die een toets heeft beantwoord, en anderzijds tussen personen, onafhankelijk van de toevallige groep items waaruit een toets heeft bestaan. De betrouwbaarheid van een toets kan zo op voorhand al worden vastgesteld omdat alleen items in de toets worden opgenomen waarvan de betrouwbaarheid hoog genoeg is (van der Linden, 1978). Dat heeft tot consequentie dat alleen items kunnen worden gebruikt die eerder in toetsen zijn opgenomen, waarvan de betrouwbaarheid bekend is en die voldoen aan de gestelde eisen.

In een vroeg stadium van het onderzoek naar de betrouwbaarheid van de huisartsgeneeskundige kennistoets is besloten dit onderzoek te baseren op de *generaliseerbaarheidstheorie*. In Bijlage 3 worden de generaliseerbaarheidstheorie en de daarin gebruikte begrippen, beschreven. Deze begrippen worden in de paragrafen hierna daarom zonder verdere toelichting gehanteerd.

Voor de keuze voor de generaliseerbaarheidstheorie gold een aantal argumenten. Eerder in kennistoetsen gebruikte items waren nog niet onderzocht op basis van itemresponsstheorieën. Vanwege de hoge eisen die deze theorie stelt aan items om in een toets te worden opgenomen (zie hiervoor), mocht worden aangenomen dat het bestand aan items te klein zou zijn om een inhoudelijk verantwoorde (dat wil zeggen volgens de blauwdruk samengestelde) toets te kunnen samenstellen. Bovendien was onbekend in hoeverre de in de huisartsgeneeskundige kennistoets gekozen vraagvorm, met meerdere items per casus, zich leent voor toepassing van deze theorieën. Bovendien worden in de analyses volgens de itemresponsstheorieën alleen dicotome of polytone scores gehanteerd. Het gebruik van goed-min-foutscores zou daarom problemen opleveren.

Een belangrijk argument voor de keuze voor de generaliseerbaarheidstheorie was verder dat deze theorie met name inzicht biedt in de factoren die de meetfout of error bepalen. Daarmee biedt deze theorie, in vergelijking met de klassieke testtheorie, het voordeel dat meerdere interpretatieperspectieven op de betrouwbaarheid van de toetsscores kunnen onderzocht (Swanson et al, 1987). Dat betekende dat de betrouwbaarheid van de toets voor verschillende gebruiksfuncties kon worden nagegaan.

3.1.1. Interpretatieperspectieven toetsscores

Toetsscores zijn getallen die van een kader moeten worden voorzien om er betekenis aan te kunnen hechten. Een toetsscore op zichzelf is zelden direct interpreteerbaar: een score van bijvoorbeeld 50 punten op een toets die maximaal 100 punten op kan leveren, kan alleen maar worden geïnterpreteerd als bekend is wat de moeilijkheidsgraad van de vragen was en/of wat de prestaties van andere toetskandidaten waren (Suen, 1990). Dit gegeven heeft onder meer tot gevolg dat er verschillende betrouwbaarheden kunnen worden berekend, afhankelijk van wat er met de toetsing wordt beoogd (Colliver et al, 1989). Als alle deelnemers aan een toets alle items maken, kunnen er drie soorten factoren worden onderscheiden die de toetsscores beïnvloeden: de kundigheid ('ability') van de toetsdeelnemers, de moeilijkheid van de items, en externe factoren (vaak interactie-effect genoemd). Idealiter reflecteren de toetsscores uitsluitend de kundigheid van de deelnemers. In psychometrische termen: de waargenomen score valt samen met de 'ware' of 'bedoelde' score. In werkelijkheid echter worden toetsscores altijd beïnvloed door de moeilijkheid van de items en door de externe factoren of interactie-effecten (Norcini & Swanson, 1989). In psychometrische termen: de waargenomen score is de resultante van de 'ware' score en de errorscore.

Zoals eerder beschreven kunnen, afhankelijk van het doel van de toetsing, het object van meting en de factoren die tot de error worden gerekend, verschillen. Volgens de generaliseerbaarheidstheorie moet vooraf worden bepaald wat het object van meting is en welke foutenbronnen wel of niet tot de errorcomponent worden gerekend. Dit levert de volgende mogelijkheden op om de toetsscores te kunnen interpreteren (Norcini & Swanson, 1989; van der Vleuten & Wijnen, 1991):

a. het *normgeoriënteerde (relatieve)* perspectief.

In het normgeoriënteerde of relatieve perspectief moeten toetsscores vooral informatie opleveren over de reproduceerbaarheid van de onderlinge rangorde van de deelnemers (de beste individuen of groepen scoren bij herhaling van de toets het hoogst, de slechtsten het laagst). Het gaat in dit perspectief om de betrouwbaarheid van de relatieve positie van de ene score ten opzichte van de andere. Alleen de verschillen in gemiddelde score tussen deelnemers of groepen van deelnemers vormen daarom het object van meting. Als de items toevallig moeilijk of gemakkelijk zijn, heeft dit alleen effect op het aantal goede of foute antwoorden van de deelnemers maar niet op hun onderlinge rangorde (de besten blijven de besten en de slechtsten blijven de slechtsten). Verschillen in moeilijkheidsgraad tussen de items vormen daarom vanuit dit perspectief geen bron van error. De betrouwbaarheid wordt uitgedrukt in een *generaliseerbaarheidscoefficiënt* (G-coëfficiënt).

b. het *domeingeoriënteerde (absolute)* perspectief.

Als een absolute betekenis aan de scores wordt toegekend (een score van 50% betekent beheersing van 50% van het getoetste domein, dat wil zeggen bij herhaling van de toets worden door dezelfde personen dezelfde scores behaald), speelt de moeilijkheidsgraad van de items wel een rol. Het gaat er dan namelijk om in hoeverre de toetsscores informatief zijn voor de reproduceerbaarheid van de mate van beheersing van de getoetste stof. De toevallige omstandigheden blijven een bron van error, maar het doet er nu ook toe of de verzameling items moeilijk of gemakkelijk was. De scores zeggen immers iets over de mate van beheersing van een bepaalde hoeveelheid stof. Het object van meting is hetzelfde als bij het relatieve perspectief, maar in de berekening van de error moeten de verschillen in moeilijkheidsgraad tussen de items nu wel worden opgenomen. De betrouwbaarheid wordt uitgedrukt in een zogenaamde *phi-coëfficiënt*.

c. het *beslissingsgeoriënteerde ('mastery oriented')* perspectief.

Het relatieve en absolute betrouwbaarheidsperspectief hebben betrekking op de betrouwbaarheid van de toetsscores. Het derde perspectief heeft betrekking op de betrouwbaarheid van beslissingen die worden genomen op basis van de toetsscores. Dit perspectief is het zogenaamde beslissingsgeoriënteerde betrouwbaarheidsperspectief ('mastery oriented perspective'). Om betrouwbaarheidsschattingen te kunnen doen vanuit dit perspectief, moet op voorhand een bepaalde grenswaarde zijn vastgelegd. Het beslissingsgeoriënteerde perspectief kan ook worden omschreven als domeingeoriënteerd (absoluut) waarbij een bepaalde grenswaarde voor zakken of slagen wordt gehanteerd (Brennan & Kane, 1977). De betrouwbaarheidsindices geven vervolgens aan hoe betrouwbaar de beslissingen zijn die op grond van deze grenswaarde worden genomen. Scores boven de grenswaarde betekenen dat de stof wordt beheerst; scores onder de grenswaarde wijzen op onvoldoende beheersing. Dat wil zeggen dat omvang van het effect van de verschillende foutenbronnen op de betrouwbaarheid, afhankelijk is van de plaats van de cesuur. In het object van meting wordt daarom ook de afstand opgenomen van de scores tot de vastgestelde grenswaarde. De betrouwbaarheid wordt uitgedrukt in een zogenaamde *adjusted phi-coëfficiënt*.

De betrouwbaarheid vanuit het beslissingsperspectief is met name van belang als men op basis van de toetsscores beslissingen wil nemen over individuen. Het verschil tussen het absolute en het beslissingsgeoriënteerde betrouwbaarheidsperspectief is gradueel. In beide gevallen gaat het om de beheersing van de stof. In het laatste perspectief gaat het er daarnaast nog om uitspraken te doen over 'voldoende beheersing' of 'onvoldoende beheersing'. Absolute

interpretaties zijn te vergelijken met verspringen. Vanaf de afzet op de balk wordt gemeten waar iemand neerkomt. Beslissingsgeoriënteerde interpretaties zijn te vergelijken met hoogspringen. Het gaat er daarbij om of iemand al dan niet over een van te voren bepaalde hoogte komt. Hoeveel eronder of erboven doet er niet toe.

3.1.2. Standaardmeetfout

De betrouwbaarheidscoëfficiënten die voortkomen uit de generaliseerbaarheidsanalyses zijn te interpreteren als correlatiecoëfficiënten die een indicatie geven van de *herhaalbaarheid* (*reproduceerbaarheid*) van de toetsscores: de kans wordt geschat in hoeverre bij herhaling van dezelfde (of een parallel-)toets bij dezelfde personen onder soortgelijke omstandigheden, hetzelfde resultaat zou worden geboekt. De coëfficiënt geeft dus een maat voor de samenhang tussen de ware, c.q. bedoelde, en de waargenomen scores, die worden aangetroffen bij de deelnemers aan de toets. In het algemeen wordt een waarde van .80 of .85 aangehouden als (onder)grens voor deze coëfficiënt, als het gaat om de betrouwbaarheid van individuele scores voor *zak/slaag*beslissingen (Carmines & Zeller, 1985; Eggen & Sanders, 1993). De *nauwkeurigheid* van een waargenomen score is een verwante, maar andersoortige indicator voor de betrouwbaarheid van toetsscores. Het betreft daarbij de vraag in hoeverre een bepaalde (waargenomen) score een goede schatting is van de ware score (Bender, 1994). Dit kan zichtbaar gemaakt worden met het betrouwbaarheidsinterval van de waargenomen score. Het betrouwbaarheidsinterval kan worden bepaald met behulp van de *Standard Error of Measurement (SEM)*, ofwel de *standaardmeetfout* (Suen, 1990; van der Vleuten et al, 1991). De standaardmeetfout is de standaarddeviatie van de meetfout (de error). Volgens de klassieke testtheorie kan deze worden berekend met behulp van de betrouwbaarheidscoëfficiënt. Volgens de generaliseerbaarheidstheorie moet eerst de error worden gedefinieerd en kan daaruit (door de wortel te nemen) de SEM worden bepaald. Volgens de generaliseerbaarheidstheorie is het dus niet nodig eerst betrouwbaarheidscoëfficiënten te berekenen (Brennan, 1983; Feldt & Brennan, 1989). De SEM ligt op dezelfde schaal als de scores en kan worden gebruikt om het gebied te bepalen waarin de ware score ligt, bij een bepaalde waargenomen score (Dousma & Horsten, 1989). Hoe kleiner de SEM is, hoe dichter de waargenomen scores bij de ware scores liggen (Meerling, 1981). Een waargenomen score plus of min éénmaal de SEM geeft het 67%-betrouwbaarheidsinterval, een waargenomen score plus of min 1.96 maal de SEM geeft het 95%-betrouwbaarheidsinterval aan waarin de ware score ligt.

In de betrouwbaarheidsstudie van de algemene huisartsgeneeskundige kennistoets is rekening gehouden met de hiervoor beschreven theoretische uitgangspunten. Hierna wordt uiteengezet hoe en met welk resultaat de studie is uitgevoerd.

Onderzoeksvraag 1.2.:

Wat is de psychometrische kwaliteit van een algemene huisartsgeneeskundige kennistoets in termen van betrouwbaarheid?

Opzet betrouwbaarheidsstudie:

- (multivariate) generaliseerbaarheidsanalyse blauwdruk: empirische inhoudsvalidering: zie paragraaf 2.3)
- betrouwbaarheid van individuele toetsscores
- betrouwbaarheid van gemiddelde toetsscores

3.2. ONDERZOEKSOPZET BETROUWBAARHEIDSSTUDIE

De betrouwbaarheidsstudie werd gericht op drie verschillende 'objecten van meting':

- een multivariate generaliseerbaarheidsanalyse werd uitgevoerd om de invloed van de *inhoudelijke structuur* (c.q. de stratificering van de toets in aantallen items per onderwerp volgens de blauwdruk) van de toets op de betrouwbaarheid vast te stellen; Het resultaat van deze analyse is beschreven in hoofdstuk 2, vanwege de eventuele consequenties voor de inhoudsvaliditeit van de kennistoets. Het resultaat van de studie was zodanig dat kon worden geconcludeerd dat een andere verdeling van items over de onderwerpen de totale toetsbetrouwbaarheid niet zou beïnvloeden. Dat wil zeggen dat er geen onderwerp uitsprong dat op de betrouwbaarheid een geheel andere invloed uitoefende dan andere onderwerpen. Tevens betekende deze bevinding dat de variantiebron 'onderwerpen' in de designs van de overige betrouwbaarheidsstudies achterwege kon worden gelaten.
- de betrouwbaarheid van *individuele scores*.
Informatie over de betrouwbaarheid van de toets op individueel niveau is noodzakelijk om uitspraken te kunnen doen over de grenzen van de bruikbaarheid van de toets als educatief hulpmiddel om sterke en zwakke kanten in de kennis van individuen te diagnosticeren. Kennisverschillen tussen personen vormen het object van meting. Zowel betrouwbaarheidscoëfficiënten (generaliseerbaarheids-, phi-coëfficiënten en adjusted phi-coëfficiënten) als standaardmeetfouten zijn bepaald voor de onderscheiden perspectieven op de interpretatie van toetsscores (Brennan, 1983).
- de betrouwbaarheid van *gemiddelde scores* (groepsscores);
Informatie over de betrouwbaarheid van gemiddelde scores is noodzakelijk om uitspraken te kunnen doen over de bruikbaarheid van de toets als screeningsinstrument (Kane & Brennan, 1977). Om betrouwbaarheidscoëfficiënten te kunnen bepalen, zou het verschil in gemiddelde kennis tussen groepen huisartsen het object van meting moeten vormen. Een probleem hierbij was dat er geen 'natuurlijk' criterium beschikbaar was om de groep huisartsen in te delen. De totale groep huisartsen zou in beginsel op talloze manieren in te delen zijn, bijvoorbeeld naar ervaring, naar leeftijd of naar soort van praktijk. De betrouwbaarheid van de groepsscores varieert met de soort groep en de variantie binnen die groep. Elke keuze voor een indeling naar groep zou daarom een element van willekeur met zich meebrengen. De oplossing die voor dit probleem is gezocht, wordt beschreven in paragraaf 3.2.2.

De formules waarmee de verschillende betrouwbaarheidscoëfficiënten en de SEM worden berekend, zijn in Bijlage 4 opgenomen. Als object van meting kan in de betrouwbaarheidsanalyses zowel voor de variantiebron 'personen' als 'groepen' worden gekozen, afhankelijk van de vraagstelling (Brennan, 1983).

Conform de gebruikelijke procedures in een generaliseerbaarheidsstudie (Cronbach et al, 1972; Brennan, 1983) werden eerst G-studies verricht waarin de verschillende variantiebronnen werden gedefinieerd. Op basis van een variantie-analyse (ANOVA) werd de relatieve bijdrage van elke variantiebron aan de totale variantie uitgedrukt in een variantie-component. Vervolgens werden, weer voor elke invalshoek afzonderlijk (toetsstructuur, individu en groep), in D-studies schattingen gemaakt van de te bereiken betrouwbaarheid bij verandering van toetssamenstelling, resp. wijziging van groepsgrootte of toetslengte.

De analyses werden uitgevoerd op de kennistoetsgegevens die in februari 1992 zijn verzameld bij ervaren huisartsen. Alle berekeningen zijn, conform de gebruikelijke procedures bij de toetsverwerking en itemanalyse in de huisartsopleiding, gebaseerd op de goed-min-foutscores, uitgedrukt in percentages van de maximaal te bereiken score, die varieert tussen -100 en +100 (Düsmen, 1992). Ook de standaardmeetfouten (SEM's) werden in deze percentageschaal uitgedrukt. In de itemanalyse werden de percentages goede, foute en met een vraagteken beantwoorde vragen berekend, en werden item-testcorrelaties en discriminatie-indices berekend om de kwaliteit van de individuele vragen in de toets te bepalen (Bender, 1994). Op basis van inhoudelijke commentaren op toetsvragen van de deelnemers en op grond van de itemanalyse zijn na afloop 14, van de oorspronkelijk 160 vragen tellende toets, uit de berekeningen weggelaten. De betrouwbaarheidsberekeningen zijn uitgevoerd over de 146 items waaruit de definitieve toets bestond waarover feedback is gegeven.

3.2.1. Bepaling betrouwbaarheid individuele toetsscores

Het *design* voor de generaliseerbaarheidsstudie naar de betrouwbaarheid van toetsscores op individueel niveau was een zogenaamd 'one facet design' met als variantiebronnen: 'personen', 'items' en de interactie tussen personen en items, die in dit design niet kan worden onderscheiden van de overige error. Zowel 'personen' als 'items' werden beschouwd als random variabelen.

Variantiebronnen

P (personen)

I (items)

PI (interactie tussen personen en items)

Toelichting:

- personen (P): verschillen tussen personen vormen het object van meting. Idealiter zou deze factor de belangrijkste bijdrage moeten leveren aan de totale toetsbetrouwbaarheid;
- items (I): deze factor geeft de verschillen in moeilijkheidsgraad tussen de items aan.
- personen en items (PI): dit interactie-effect geeft aan in welke mate personen anders gerangordend worden door de verschillende items. Dit effect is in dit design niet te onderscheiden van allerlei andere foutenbronnen die de score beïnvloeden.

In de *D-studies* is vervolgens op basis van de in de *G-studie* berekende variantiecomponenten, de invloed geschat van verandering van het aantal items op de te bereiken betrouwbaarheid. De berekeningen zijn zowel vanuit relatief, als vanuit absoluut en beslissingsgeoriënteerd betrouwbaarheidsperspectief gemaakt. Bovendien is berekend wat de standaardmeetfouten zijn bij de verschillende betrouwbaarheidscoëfficiënten.

3.2.2. Bepaling betrouwbaarheid groepsscores

Het *design* voor de studie naar de betrouwbaarheid van gemiddelde toetsscores van groepen huisartsen is in termen van de generaliseerbaarheidstheorie te beschrijven als een 'two facet design': groep x personen (binnen groep) x items (Brennan, 1983). 'Groep' is hierbij het object van meting; de variantie ten gevolge van 'groep' is dus de bedoelde variantie. 'Personen binnen groepen' en 'items' zijn facetten. Deze facetten zijn random variabelen. Een

relevant interactie-effect in het design vormt het interactie-effect van 'groep' met 'items'. Omdat elke persoon zich maar in één groep kan bevinden (in generaliseerbaarheidstermen zijn de personen 'genest' binnen groepen), is de variantiebron 'personen' niet te onderscheiden van het interactie-effect tussen 'personen' en 'groepen'. De variantie ten gevolge van interactie tussen 'personen' en 'items' is dus in dit design ook niet als separate variantiebron te onderscheiden en vormt onderdeel van de totale errorvariantie. Dat betekent dat de volgende *variantiebronnen* zijn onderscheiden.

Variantiebronnen

G	(groep)
I	(items)
P:G	(personen genest binnen groepen)
GI	(interactie groep met items)
PI:G	(variantie interactie plus overige variantie)

Toelichting:

- groepen (G): de verschillen in gemiddelde kennis van groepen huisartsen vormen hierbij het object van meting;
- items (I): deze variantiebron betreft de bijdrage aan de variantie van verschillen in moeilijkheidsgraad tussen items;
- personen binnen groepen (P:G): de mate van verschillen tussen personen binnen groepen: de binnengroepsvariantie;
- groepen en items (GI): hiermee wordt de variantie uitgedrukt die teweeg gebracht wordt door het interactie-effect van groepen met items. Het geeft aan in welke mate bepaalde groepen op bepaalde items anders scoren dan andere groepen;
- de variantie van personen met items (binnen groepen) (uitgedrukt als PI:G). In dit interactie-effect is alle overige errorvariantie opgenomen.

Dit design is van toepassing als de betrouwbaarheid van de scores van bepaalde bestaande groepen met elkaar worden vergeleken. In de onderhavige studie ging het echter om het vinden van een goede indicator voor de betrouwbaarheid van de algemene kennistoets als screeningsinstrument. Het gaat er daarbij met name om bij welke groeps grootte een acceptabele betrouwbaarheid wordt bereikt en niet zozeer om bijvoorbeeld de betrouwbaarheid van de rangordening van de gemiddelde scores van verschillende gespecificeerde subgroepen. Dat betekent dat de gegevens zich er niet toe leenden om generaliseerbaarheidscoëfficiënten op groepsniveau te bepalen. Daarom is de *standaardmeetfout* (SEM) op groepsniveau gekozen als belangrijkste indicator voor de groepsbetrouwbaarheid. Binnen de generaliseerbaarheidstheorie is het mogelijk de standaardmeetfout van groeps gemiddelden te berekenen als deze bekend is voor individuele scores (Brennan, 1983). In de berekeningen kan worden volstaan met de variantiecomponenten P, I en PI, uit het design voor de bepaling van de betrouwbaarheid van individuele toetsscores (Brennan, 1983). Voor de formules wordt verwezen naar Bijlage 4. Standaardmeetfouten werden berekend bij variabele groeps grootte en toetslengte. De SEM's werden uitgedrukt in dezelfde percentuele schaal eenheid als de gemiddelde goed-min-fout scores die als groepsscores werden gehanteerd.

3.3. RESULTATEN BETROUWBAARHEIDSSSTUDIE

3.3.1. Betrouwbaarheid individuele toetsscores

In tabel 3.1 worden de variantiecomponenten weergegeven, die zijn onderscheiden in de analyses naar de generaliseerbaarheid van de scores op individueel niveau, inclusief hun bijdrage aan de totale variantie. De gemiddelde goed-min-foutscore van de huisartsen was 45 op een schaal die loopt van -100 (alle items fout) tot +100 (alle items goed).

Tabel 3.1. *Geschatte variantiecomponenten en percentage van de totale variantie van kennistoetsscores van huisartsen (N=364).*

Variantiebron	Geschatte variantiecomponenten	Percentage van de totale variantie
P (N=364)	.009008	1.3
I (N=146)	.150221	20.6
PI	.569746	79.1

Uit tabel 3.1 blijkt dat de bijdrage aan de totale variantie van de variantiebron 'personen' in vergelijking met de variantie die samenhangt met 'items' klein is. Deze bevinding wordt ook gevonden in de huisartsopleiding (Düsman, 1992; van Leeuwen et al, in voorbereiding).

Tabel 3.2 toont de schattingen van de betrouwbaarheidsindices ook bij verandering van het aantal items.

Tabel 3.2. *Betrouwbaarheid (uitgedrukt in generaliseerbaarheidscoëfficiënten, phi-coëfficiënten en SEM) van de algemene kennistoets voor individuele huisartsen (N=364) bij wisselende aantallen items.*

Aantal items	Toetsduur (uren)	Generaliseerbaarheids-coëfficiënt (relatief)	SEM	Phi-coëfficiënt (absoluut)	SEM
20	0.25	.24	17	.20	19
40	0.5	.39	12	.33	13
80	1	.56	8	.50	9
146*	2	.70	6	.65	7
160	2	.72	6	.67	7
240	3	.79	5	.75	6
320	4	.84	4	.80	5

* feitelijke toetslengte

De betrouwbaarheid van de totale toets (berekend over de 146 vragen die resteerden nadat de items waarmee achteraf inhoudelijke problemen bleken te bestaan, waren verwijderd) voor de 364 huisartsen waarvoor goed-min-foutscores konden worden berekend, bedroeg .70 vanuit het relatieve perspectief (met een SEM van 6) en .65 vanuit het absolute perspectief (met een SEM van 7). Voor de interpretatie van een gemiddelde score (45%) betekent dit dat de 'ware score' (vanuit relatief perspectief) met 67% zekerheid ligt tussen 39 en 51%, en met 95% zekerheid tussen 33 en 57%. Vanuit absoluut perspectief zijn deze percentages 38 en 52 (67% zekerheid), respectievelijk 31 en 59 (95% zekerheid).

De betrouwbaarheid van de individuele scores komt bij een toetsduur van ongeveer 3 uur (240 items) in de buurt van de .80 (relatief perspectief). Voor het absolute perspectief, waarbij de scores de kennisbeheersing uitdrukken, geldt dit voor een toetsduur van ongeveer 4 uur. De bijbehorende SEM is in beide gevallen 5. Met name dit laatste gegeven maakt duidelijk wat de toename is van de nauwkeurigheid waarmee wordt gemeten bij verlenging van de toets. In het gegeven voorbeeld (voor een gemiddelde score van 45%) zou uitbreiding van de toets tot 240 of 320 items, dat wil zeggen een toetsduur van 3, respectievelijk 4 uur, de marges met ongeveer eenderde verkleinen. Dat betekent dat de meetfout bij een toets met een betrouwbaarheid van .80 niet veel kleiner is dan in de bestaande situatie met een toetsbetrouwbaarheid van .70 (zie tabel 3.3).

Tabel 3.3. Betrouwbaarheidsintervallen van gemiddelde individuele toetsscore van huisartsen (gemiddelde = 45% goed-min-fout) bij verschillende toetslengtes en verschillende betrouwbaarheidsperspectieven.

Toetslengte	Betrouwbaarheidsinterval gemiddelde individuele toetsscore (45%)			
	Relatief		Absoluut	
	67%-interval	95%-interval	67%-interval	95%-interval
160 items (2 uur)	39-51	33-57	38-52	31-59
240 items (3 uur)	40-50	35-55	39-51	33-57
320 items (4 uur)	41-49	37-53	40-50	35-55

Tabel 3.4 levert het resultaat van de betrouwbaarheidsberekeningen voor individuele scores vanuit het beslissingsperspectief bij verschillende grensscores en toetslengtes.

Tabel 3.4. Betrouwbaarheid (uitgedrukt in adjusted phi-coëfficiënten) van de algemene kennistoets voor individuele huisartsen vanuit beslissingsgeoriënteerd perspectief.

Aantal items	Toets-tijd	Betrouwbaarheid bij verschillende grensscores								
		20	30	35	40	45*	50	55	60	70
10	10'	.44	.18	.05	0	0	0	.05	.18	.44
20	15'	.64	.40	.24	.10	.04	.10	.24	.40	.64
30	20'	.74	.52	.36	.21	.14	.21	.37	.53	.74
40	30'	.79	.60	.45	.30	.22	.30	.46	.61	.79
80	1 uur	.88	.76	.65	.51	.44	.52	.66	.77	.89
146**	2 uur	.93	.86	.78	.68	.62	.68	.79	.86	.93
160	2 uur	.94	.87	.80	.70	.64	.70	.80	.87	.94

*: gemiddelde toetsscore

** : feitelijke toetslengte

De tabel laat zien dat naarmate de grensscore verder onder of boven de gemiddelde totaalscore (45%) wordt gelegd, de betrouwbaarheid van de beslissingen toeneemt. Voor de complete toets van 160 vragen zijn betrouwbare beslissingen mogelijk bij een grenswaarde

van 35% of minder (beslissing is dan: 'kennis wordt onvoldoende beheerst') of van meer dan 55% (beslissing: 'kennis wordt voldoende beheerst').

3.3.2. Betrouwbaarheid groepsscores

In de weergave van de resultaten wordt de betrouwbaarheid van de groepsgemiddelden aangegeven met de standaardmeetfout die werd berekend op basis van de hiervoor gerapporteerde individuele betrouwbaarheidsgegevens (voor de formule waarmee de gegevens zijn berekend wordt verwezen naar Bijlage 4).

Tevens is geschat welke verandering wordt teweeggebracht bij wijziging van groepsgrootte en toetslengte. Het resultaat is opgenomen in de volgende tabellen. In tabel 3.5 worden de SEM's gepresenteerd van de gemiddelde scores vanuit relatief perspectief.

Tabel 3.5. Betrouwbaarheid (uitgedrukt in SEM) van gemiddelde groepsscores van huisartsen op de algemene kennistoets vanuit relatief perspectief.

Aantal items	Toets-tijd	Standaardmeetfout (SEM) bij verschillende groepsgroottes						
		N=10	N=20	N=30	N=50	N=100	N=200	N=365
10	10'	8	6	5	4	3	2	1
20	15'	6	4	4	3	2	1	1
30	20'	5	4	3	2	2	1	1
40	30'	4	3	3	2	2	1	1
80	1 uur	4	3	2	2	1	1	1
160	2 uur	4	3	2	1	1	1	1

De tabel maakt duidelijk dat de precisie van de gemiddelde toetsscores vanuit het relatieve perspectief bij een groepsgrootte van 20 tot 30 personen hoog is (een geschatte afwijking van ongeveer 2%). Betrouwbare groepsscores worden dus al bereikt bij een beperkt aantal deelnemers aan de toets. Bij deelname van 50 tot 100 personen is de betrouwbaarheid van de gemiddelde groepsscores maximaal.

In tabel 3.6 is de betrouwbaarheid van de gemiddelde groepsscores bij verschillende groepsgrootte en toetslengtes weergegeven vanuit het absolute perspectief.

Tabel 3.6. Betrouwbaarheid (uitgedrukt in SEM) van gemiddelde groepsscores van huisartsen op de algemene kennistoets vanuit absoluut perspectief.

Aantal items	Toets-tijd	Standaardmeetfout (SEM) bij verschillende groepsgroottes						
		N=10	N=20	N=30	N=50	N=100	N=200	N=365
10	10'	15	14	13	13	12	12	12
20	15'	11	10	9	9	9	9	9
30	20'	9	7	7	7	7	7	7
40	30'	8	7	7	6	6	6	6
80	1 uur	6	5	5	5	4	4	4
160	2 uur	5	4	3	3	3	3	3

Bij de berekening van de betrouwbaarheidsindices vanuit absoluut perspectief, is de variantie ten gevolge van de moeilijkheidsgraad van de items betrokken (zie Bijlage 4 voor de formule). Daarom vallen de betrouwbaarheidsschattingen lager uit dan volgens het relatieve perspectief. De precisie van de gemiddelde groepsscores vanuit absoluut betrouwbaarheids perspectief blijkt ook hier groter is naarmate de toetsen langer zijn en/of het aantal deelnemers groter is.

Ter illustratie van de in de vorige tabellen gepresenteerde cijfers, is in tabel 3.7 weergegeven met welke betrouwbaarheidsintervallen rekening moet worden gehouden bij de interpretatie van toetsscores van een groep huisartsen vanuit het relatieve en vanuit het absolute perspectief. Als uitgangspunt is de gemiddelde toetsscore van de huisartsen genomen (45% goed-min-fout) en is gekozen voor een groep met een omvang van 30 huisartsen.

Tabel 3.7. *Betrouwbaarheidsintervallen van gemiddelde groepsscore van huisartsen (gemiddelde goed-min-foutscore=45%) bij verschillende toetslengtes en verschillende betrouwbaarheidsperspectieven. (groep van 30 huisartsen als object van meting)*

Toetslengte	Betrouwbaarheidsinterval gemiddelde groepsscore (45%; N=30)			
	Relatief		Absoluut	
	67%-interval	95%-interval	67%-interval	95%-interval
20 items (15')	41-49	37-53	36-54	27-63
40 items (30')	42-48	39-51	38-52	31-57
80 items (1 uur)	43-47	41-49	40-50	35-55
160 items (2 uur)	43-47	41-49	42-48	39-51

Een gemiddelde goed-min-fout-groepsscore van 45% voor een groep van 30 huisartsen op een toets van 20 items, betekent vanuit relatief perspectief dat er 67% kans bestaat dat de ware gemiddelde groepsscore ligt tussen 41 en 49%, en dat er 95% kans bestaat dat de ware gemiddelde groepsscore ligt tussen 37 en 53%. Als wordt uitgegaan van een absoluut perspectief moet met ruimere grenzen rekening worden gehouden (36 tot 54%; resp. 27 tot 63%). Hoe langer de toetsen zijn, des te nauwkeuriger de groepsgemiddelden. Voor de complete toets van 160 vragen moet men uitgaan van een standaardmeetfout (vanuit relatief perspectief) van 2%. Dat wil zeggen dat er 95% kans bestaat dat de ware gemiddelde groepsscore van deze groep huisartsen tussen de 41 en 49% ligt. Als wordt uitgegaan van een absoluut perspectief moet rekening worden gehouden met een SEM van 3. Dat betekent dat ervan mag worden uitgegaan dat er een kans bestaat van 95% dat de ware beheersing bij de groep huisartsen van de getoetste stof ligt tussen 39 en 51%.

3.4. CONCLUSIES BETROUWBAARHEIDSTUDIE ALGEMENE HUISARTSGENEESKUNDIGE KENNISTOETS

De studie naar de betrouwbaarheid van de algemene huisartsgeneeskundige kennistoets is gebaseerd op de generaliseerbaarheidstheorie. Voor de betrouwbaarheid van de individuele toetsscores zijn op basis van deze theorie verschillende betrouwbaarheidscoëfficiënten

berekend, overeenkomend met de perspectieven op de betrouwbaarheid die bij verschillende gebruiksfuncties van de toets in acht moeten worden genomen.

Op individueel niveau geeft de betrouwbaarheid een indruk van de reproduceerbaarheid en de nauwkeurigheid van de individuele score. De betrouwbaarheid van individuele scores van huisartsen is .70 vanuit het relatieve perspectief en .65 vanuit absoluut perspectief. Daarbij horen foutenmarges (uitgedrukt in de SEM) van plus of min 6 (67%-betrouwbaarheidsinterval) of 12 (95%-betrouwbaarheidsinterval) voor de betrouwbaarheid van de relatieve rangordening, en marges van plus of min 7, resp. 14 voor de betrouwbaarheid van de toetsscores voor wat betreft de beheersing van de getoetste stof. Deze marges lijken te groot om aan de scores individuele consequenties te verbinden. Bij een individuele score van bijvoorbeeld 45% (de gemiddelde goed-min-foutscore van de huisartsen) moet rekening worden gehouden met een kans van 67% op een ware score die tussen 38 en 52 ligt, en met een waarschijnlijkheid van 95% dat de score ligt tussen 31 en 59. Uit de berekeningen bleek dat verlenging van de toetsen wel kan leiden tot hogere betrouwbaarheidscoëfficiënten, maar slechts een marginaal effect heeft op de nauwkeurigheid van de meetgegevens. Verlenging van de toets tot twee maal de huidige omvang (hetgeen een toetstijd betekent van drie tot vier uur), betekent dat het betrouwbaarheidsinterval 1 schaal eenheid kleiner wordt. De veel grotere inspanning die zowel toetsconstructeurs als huisartsen die deelnemen aan de toets moeten leveren, levert dus maar marginale nauwkeurigheidswinst op. Deze gegevens gelden de totale toetsbetrouwbaarheid. De betrouwbaarheid van individuele scores op onderdelen van de toets is nog lager. Voor uitspraken over het niveau van kennisbeheersing van individuele huisartsen op onderdelen van de toets, leent de algemene kennistoets zich dan ook minder.

De betrouwbaarheid vanuit het beslissingsgeoriënteerde perspectief, dat wil zeggen de betrouwbaarheid van beslissingen, is voor individuele scores acceptabel als de grenswaarden maar niet te dicht worden gekozen bij de gemiddelde score. Dat betekent dat de toets op individueel niveau in staat is om betrouwbare beslissingen op de toetsscores te baseren onder de voorwaarde dat er overeenstemming is te vinden over een acceptabele grensscore. Dit is echter geen sinecure. Het is immers onduidelijk welk minimumniveau aan kennis verenigbaar is met het goed functioneren als huisarts.

De huisartsen van wie de toetsscores zijn gebruikt voor de betrouwbaarheidsschattingen, lieten zich op voorhand niet indelen in 'natuurlijke' subgroepen waar scoreverschillen waarover betrouwbaarheidscoëfficiënten zouden kunnen worden berekend, tussen mochten worden verwacht. Voor de groepsgemiddelden werd daarom als betrouwbaarheidsindicator de standaardmeetfout (SEM) gehanteerd die werd berekend op basis van de individuele gegevens. Op groepsniveau drukt de betrouwbaarheid uit welke foutenmarges in acht genomen moeten worden bij de interpretatie van een bepaald groepsgemiddelde. Dit is uiteraard vooral van belang als een bepaald groepsgemiddelde afwijkt van het totaal gemiddelde. De kennistoets levert scores op die betrouwbaar genoeg zijn om uitspraken te doen over verschillen in kennisniveau tussen groepen van huisartsen. De scores (op de totale toets) bereiken een maximale nauwkeurigheid (standaardmeetfout 1 vanuit het relatieve perspectief) voor groepen die bestaan uit 30 tot 50 huisartsen. Als de betrouwbaarheid wordt beschouwd vanuit het absolute perspectief, is de standaardmeetfout voor groepen met een dergelijke omvang 3. Vergroting van de groep deelnemers maakt de standaardmeetfout niet wezenlijk kleiner. De gemiddelde scores zijn op onderdelen van de toets (van 20 tot 30 vragen) die door relatief kleine groepen huisartsen worden gemaakt, nog steeds nauwkeurig, zeker vanuit het relatieve perspectief.

Deze conclusies laten onverlet dat de toets een belangrijke educatieve functie kan hebben. Zowel op groepsniveau als op individueel niveau kan feedback over de toetsresultaten aan de deelnemers worden gegeven. Omdat de betrouwbaarheidsstudie inzicht geeft in de marges waarmee bij de interpretatie van deze scores rekening moet worden gehouden, betekent dit dat huisartsen een goed inzicht kan worden gegeven in de relatieve betekenis van de behaalde scores.

De resultaten van de betrouwbaarheidsstudie kunnen op de volgende manier in een breder perspectief worden geplaatst. De betrouwbaarheid van toetsresultaten hangt nauw samen met de toetslengte (Levine et al, 1970; Swanson et al, 1991; van der Vleuten, 1991). Dit heeft te maken met de inhoudsspecificiteit van probleemoplossende vaardigheden, die ertoe leidt dat veel verschillende problemen (in kennistoetsen geoperationaliseerd als vragen) nodig zijn om een betrouwbaar beeld van de kennis van de deelnemers aan de toets te verkrijgen (van der Vleuten & Newble, 1994). Vergelijking van de gegevens in de onderhavige studie, met andere onderzoeksgegevens over de betrouwbaarheid van schriftelijke meerkeuzetoetsen in het medisch domein, laat geen grote verschillen zien, als met de toetslengte rekening wordt gehouden. Norcini et al (1984) rapporteren betrouwbaarheidscoëfficiënten van .76 (toetstijd 1.7 uur) tot .83 (toetstijd 2.3 uur) voor meerkeuzetoetsen, gebruikt in certificatie-examens op het gebied van de interne geneeskunde, waaraan bijna 8000 kandidaten deelnamen. De betrouwbaarheid (uitgedrukt als Cronbach's alpha) van de goedscores, die door huisartsen-in-opleiding werden behaald op een landelijke kennistoets bestaande uit 115 juist-onjuist-vraagtekenvragen over diabetes mellitus, bedroeg .79 (Zuidweg, 1994). De betrouwbaarheid van goed-min-foutscores op deze toets is niet gerapporteerd, maar zal gezien de hoogte van de betrouwbaarheid van de goedscores, ongeveer gelijk zijn aan de betrouwbaarheid die in het onderhavige onderzoek voor de algemene huisartsgeneeskundige kennistoets werd gevonden.

De Maastrichtse voortgangstoets is het instrument dat het meest vergelijkbaar is met de algemene huisartsgeneeskundige kennistoets (Wijnen, 1986; van der Vleuten & Verwijnen, 1990). In het laatste jaar van de universitaire artsopleiding wordt op deze toets, die gemiddeld uit ongeveer 250 tot 300 juist-onjuist-vraagtekenvragen bestaat, een gemiddelde goed-min-foutscore van ongeveer 40% behaald (van Hessen e.a., 1991). De betrouwbaarheid van de goed-min-foutscores die op deze toets wordt bereikt, is in het algemeen hoog: ongeveer .85 tot .90 (van der Vleuten, 1985; Verwijnen et al, 1987; van der Vleuten & Verwijnen, 1990). Deze betrouwbaarheid wordt echter berekend voor de totale groep deelnemers aan de voortgangstoetsen, dat wil zeggen studenten uit het universitaire medische curriculum van jaar 1 tot en met jaar 6. De betrouwbaarheid van een toets is echter in het algemeen hoger naarmate de groep deelnemers meer heterogeen is (Drenth & Sijtsma, 1990). Voor de afzonderlijke jaargroepen zijn de betrouwbaarheidscijfers dan ook lager (Imbos, 1989; van der Vleuten & Verwijnen, 1990).

Een kanttekening die bij de gerapporteerde betrouwbaarheid voor de algemene kennistoets moet worden gemaakt, is dat de betrouwbaarheidsschattingen zijn gebaseerd op een 'conservatieve' benadering van de betrouwbaarheidsproblematiek. De berekeningen zijn gebaseerd op goed-min-foutscore. Dit is de score waarbij gecorrigeerd is voor eventueel gokgedrag door in de scoreberekening foute antwoorden in mindering te brengen op het aantal correcte antwoorden. Schattingen van de betrouwbaarheid, gebaseerd op de zogenaamde 'goedscore', waarin niet voor gokken is gecorrigeerd, levert hogere betrouwbaarheden op. Een

andere kanttekening is dat de op basis van de generaliseerbaarheidstheorie berekende betrouwbaarheidscoëfficiënten vanuit het relatieve perspectief, gelijk zijn aan de zogenaamde Cronbach's alpha's. Cronbach's alpha is een betrouwbaarheidsmaat vanuit de klassieke testtheorie, die de interne consistentie van de toets weergeeft. Deze betrouwbaarheidsmaat moet worden beschouwd als een ondergrens voor de betrouwbaarheid (Drenth & Sijsma, 1990).

Hoofdstuk 4

PRAKTISCHE TOEPASBAARHEID ALGEMENE HUISARTSGENEESKUNDIGE KENNISTOETSING

SAMENVATTING

De praktische toepasbaarheid van algemene huisartsgeneeskundige kennistoetsing is onderzocht door het inventariseren van de mening van ervaren huisartsen die de algemene huisartsgeneeskundige kennistoets maakten, over de kwaliteit van de toets en de toetsvragen, en over de toepasbaarheid van de toets voor een aantal gebruiksdoelen. De uitvoerbaarheid van de methode is verder nagegaan door de werkzaamheden te inventariseren die nodig zijn om een goede toets te construeren en te kunnen afnemen, door de tijd te inventariseren die gemoeid is met het invullen van een toets en door na te gaan welke activiteiten nodig zijn voor een adequate gegevensverwerking en feedbackverzorging.

De huisartsen oordelen in meerderheid positief over de toets. Bijna 90% van de huisartsen beoordeelde de vragen en hoofdstukken waaruit de toets bestond, als huisartsgeneeskundig relevant. Men vindt de toets vooral goed bruikbaar als instrument voor zelfbeoordeling. Het invullen van een toets van 160 vragen kost gemiddeld ongeveer anderhalf uur. Omdat beschikt kon worden over een bestaand systeem van vraag- en toetsproductie, gegevensverwerking en -analyse en bestaande systemen voor feedbacksamenstelling verliep de uitvoering zonder problemen.

De conclusie luidt dat het goed mogelijk is een voor huisartsen relevante kennistoets samen te stellen en af te nemen. Daarbij geldt als voorwaarde dat gebruik kan worden gemaakt van bestaande procedures voor wat betreft toetssamenstelling, gegevensverwerking en -analyse, en feedbackverzorging. Een belangrijk aandachtspunt is de kwaliteit van de toetsvragen en de expertise waarover moet worden beschikt om goede toetsvragen op een verantwoorde wijze te kunnen blijven samenstellen.

4.1. INLEIDING

Een belangrijke factor die de acceptabiliteit van een toetsmethode voor een bepaalde doelgroep beïnvloedt, is de mate waarin de toetsmethode de werkelijkheid representeert. De gelijkenis met de werkelijkheid appelleert aan een intuïtief begrip van validiteit (ofwel face validity) van de toetsmethode waarmee de acceptabiliteit direct samenhangt (van der Vleuten & Norman, 1991).

De opvattingen van de doelgroep, de ervaren huisartsen, over de algemene huisartsgeneeskundige kennistoets werden daarom nagegaan. Tevens wordt ingegaan op de praktische

uitvoerbaarheid van de procedures die nodig zijn voor het toetsen van kennis van ervaren huisartsen en het terugrapporteren van resultaten aan de deelnemers.

Onderzoeksvraag I.3.:

Wat is de praktische toepasbaarheid in termen van acceptabiliteit en uitvoerbaarheid, van een algemene kennistoets voor huisartsen?

Opzet onderzoek praktische toepasbaarheid:

- inventarisatie satisfactie huisartsen met kennistoetsing
- registratie benodigde inzet van mensen en middelen

4.2. METHODE BEPALING PRAKTISCHE TOEPASBAARHEID

De bepaling van de praktische toepasbaarheid van de kennistoets had noodzakelijkerwijs een meer beschrijvend en kwalitatief karakter, dan de bepaling van de validiteit en betrouwbaarheid van de toets in beide voorgaande hoofdstukken.

Om de acceptabiliteit van de toets voor de huisartsen te kunnen nagaan werd een satisfactie-vragenlijst, bestaande uit een beperkt aantal vijfpuitsstellingen en enkele open vragen over de inhoud en toepasbaarheid van de toets, voorgelegd aan de ervaren huisartsen die in februari 1992 deelnamen aan de op de blauwdruk gebaseerde algemene kennistoets (zie Bijlage 7). De vragenlijst bevatte vragen over:

- de huisartsgeneeskundige relevantie van de vragen en de onderwerpen (gebaseerd op de blauwdruk-categorieën) waaruit de toets bestond,
- hoe de huisartsen deelname aan de toets ervoeren,
- de lengte van de toets, en
- hoe lastig ze de vraagvorm (juist-onjuist-vraagteken) vonden.

Tevens werd gevraagd naar het oordeel van de huisartsen over de bruikbaarheid van de kennistoets:

- als hulpmiddel om eigen kennislacunes vast te stellen, en
- als uitgangspunt voor het bepalen van het eigen nascholingsprogramma.

De vragenlijst bood ruimte voor algemeen commentaar op de toets en voor specifiek commentaar op de inhoud van toetsvragen.

De inzet aan middelen en menskracht die nodig is voor de ontwikkeling van toetsmateriaal, de samenstelling en verspreiding van kennistoetsen, de organisatie van toetsafnames, de verwerking en analyse van gegevens, het samenstellen van adequate terugrapportages etc. werd gedurende de looptijd van het onderzoek bijgehouden. Hierbij is tevens gebruik gemaakt van de ervaring die inmiddels met het produceren en verwerken van kennistoetsen is opgedaan bij het landelijk Samenwerkingsverband, dat de kennistoetsen voor de huisartsopleiding produceert. De feitelijke belasting die de verschillende activiteiten met zich meebrengen, is door middel van interviews met de betrokkenen vastgesteld.

4.3. RESULTATEN

4.3.1. Acceptabiliteit

De kennistoets van februari 1992 was voorgelegd aan 365 huisartsen. Van 353 huisartsen werden ingevulde satisfactieformulieren ontvangen. In tabel 4.1 worden de oordelen en opvattingen van deze huisartsen weergegeven.

Tabel 4.1. *Opvattingen van ervaren huisartsen (N=353) over de algemene huisartsgeneeskundige kennistoets.*

Opvattingen	Waardering (in percentages)			
	(geheel) eens	neutraal	(geheel) on-eens	geen mening
Prettig toets te maken	78	18	4	—
Kennistoets is te lang	28	33	39	—
Toetshoofdstukken relevant	90	8	2	—
Toetsvragen relevant	85	11	4	—
Vraagvorm (JOV) is lastig	19	18	63	—
Toets nuttig voor zelfbeoordeling	80	13	6	2
Toets nuttig bij keuze nascholing	64	21	11	4

De meeste huisartsen apprecieerden de toets en beoordeelden zowel de hoofdstukken als de vragen van de toets als (zeer) relevant. Ongeveer een derde deel van de ervaren huisartsen vindt de toets te lang; ongeveer evenveel huisartsen zijn een tegenstelde mening toegedaan. De meeste huisartsen geven aan geen moeite te hebben met de gehanteerde juist-onjuist-vraagteken vraagvorm.

Tachtig procent van de ervaren huisartsen die aan de kennistoets deelnamen, is van mening dat een algemene kennistoets een nuttig hulpmiddel kan zijn voor het opsporen van eigen kennislacunes. Bijna twee derde deel van de deelnemende huisartsen meent dat de kennistoets nuttig is om een gerichte keuze voor bepaalde nascholings thema's te maken.

De vragenlijst bood gelegenheid algemeen commentaar te geven op de toets. De meeste opmerkingen waren positief en benadrukten de praktische relevantie van de toets voor de huisarts. In het volgende overzicht worden ter illustratie de meest genoemde opmerkingen gerubriceerd. Het overzicht geeft niet de frequentie aan waarmee opmerkingen werden gemaakt.

Tabel 4.2. Algemene opmerkingen ervaren huisartsen over huisartsgeneeskundige kennistoets.

Negatief	Positief
<p>Niveau</p> <ul style="list-style-type: none"> - opleiding tot hoogleraar - soms te detaillistisch - sommige zaken komen weinig voor - sommige onderwerpen kunnen diepgaander - niet alle vragen behoren tot parate kennis - vereist teveel recente kennis die ik niet beheers - te veel op haio's afgestemd - te weinig over continuïteit, gezinsgeneeskunde 	<p>Niveau</p> <ul style="list-style-type: none"> - sluit goed aan op de huisartspraktijk - beter dan andere toetsen - goede weergave van het vak - inhoudelijk goede toets - een gevarieerde doorsnede
<p>Vorm</p> <ul style="list-style-type: none"> - gevoel dat vragen zijn opgesteld door niet-huisartsen van achter een bureau - vermijd dat kennis van Nederlandse taal in plaats van inhoudelijke kennis wordt getoetst - schriftelijke casus zijn meestal duidelijker dan de praktijk - juist-onjuist vorm leidt ertoe dat je aan je eigen kennis gaat twijfelen - de werkelijkheid is niet zo juist-onjuist - casus vergen te veel leestijd - te veel vragen leiden tot leesmoeheid - ik zou mijn antwoorden willen motiveren 	<p>Vorm</p> <ul style="list-style-type: none"> - mijn complimenten: duidelijk gemaakt door huisartsen - goed leesbaar en begrijpelijk - prettig aan de hand van casus - slimme vragen - prettig dat veel vragen een vrouwelijke huisarts hebben - literatuurverwijzingen zijn een fantastisch idee en hulpmiddel
<p>Gebruik</p> <ul style="list-style-type: none"> - ik doe soms anders dan volgens de vraag nodig is, maar mijn aanpak is ook goed - bruikbaarheid te beperkt om lacunes voor bepaalde onderwerpen te vinden - toetst geen attitude - ik ben het soms niet eens met de auteurs van de artikelen waarnaar wordt verwezen 	<p>Gebruik</p> <ul style="list-style-type: none"> - opzoeken antwoorden achteraf is een vorm van nascholing op zich - ga toets gebruiken om gaten in kennis op te sporen - goed om lacunes in kennis te diagnosticeren - toets vervult voor solist tot op zekere hoogte de functie van een kritisch meekijkende ander - goed voor zelfreflectie - akelig duidelijk wordt hoe vluchtig kennis is: goed bruikbare toets die je aan het denken zet - stimuleert tot bijhouden literatuur - zinvol dit elk kwartaal te herhalen en in nascholingsgroep te bespreken
<p>Algemeen</p> <ul style="list-style-type: none"> - moeilijk - Standaarden zijn niet zaligmakend - ik doe mijn best zo veel mogelijk te vergeten 	<p>Algemeen</p> <ul style="list-style-type: none"> - gemakkelijk - goede kennismakingstoets - eigen kennis valt tegen - leuke leer- en ervaringstoets - confronterend en stimulerend - boeiend, leerzaam, nuttig, fijn, prima, interessant, uitdagend - graag de volgende toets: voor herhaling vatbaar

Een huisarts merkte op: 'Prima systeem! Heb maar één vraagteken ingevuld! De rest + of -, vooral zoals ik het doe (of denk te doen).'

Tabel 4.2 beoogt een globaal overzicht te geven van de gemaakte opmerkingen. Verschillende opmerkingen kwamen vaker terug. Dit gold vooral de opmerkingen die onder 'niveau' zijn gerubriceerd over de (ongewenste) detaillering van een aantal vragen, onder 'vorm' over de als lastig geïnterpreteerde juist-onjuist-vraagtekenvorm van de toetsvragen. Veel positieve opmerkingen werden gemaakt over het inhoudelijke niveau van de toets en de aansluiting op de praktijk, en over de aanwezigheid van literatuurreferenties.

4.3.2. Praktische uitvoerbaarheid

De praktische uitvoerbaarheid wordt besproken vanuit de optiek van de toetsvragenproductie, de feitelijke toetsafname en vanuit de optiek van de analyse en verwerking van de toetsgegevens.

Het *produceren van goede toetsvragen* vereist specifieke vaardigheden (Haladyna & Downing, 1989-I; 1989-II; van Susteren et al, 1991). De productie van goede toetsvragen is geen activiteit die individueel kan worden uitgevoerd, maar vergt een team van deskundige vragenmakers die onderling overleggen over de relevantie, de vorm en de inhoud van de vragen. De formulering van toetsvragen luistert nauw en beïnvloedt de kwaliteit en de bruikbaarheid van de uiteindelijke toets in hoge mate (Eekhof e.a., 1992). Ambigue vragen beïnvloeden daarbij de betrouwbaarheid van de toets in negatieve zin. De kwaliteit van het toetsmateriaal is, met andere woorden, direct van invloed op wat er met de toets wordt gemeten (Swanson et al, 1987). Controle van toetsvragen door een beoordelingscommissie van ter zake deskundigen, voorafgaand aan opname van de vragen in een toets wordt dan ook beschouwd als een essentiële procedure om de kwaliteit van de toets te verbeteren (van der Vleuten & Verwijnen, 1990; Verwijnen e.a., in voorbereiding). Het nut van zo'n commissie is niet alleen het vermijden, c.q. voorkómen van opname van verkeerde vragen in een toets, maar leidt er tevens toe dat vragenmakers hun opvattingen over de te toetsen kennis moeten expliciteren, dat keuzes worden gefundeerd, en dat 'bedrijfsblindheid' wordt geneutraliseerd of pre-occupaties worden ontmaskerd. Een centrale toetscommissie onderstreept tenslotte een gemeenschappelijke verantwoordelijkheid voor de inhoud van de toets (Wijnen, 1993).

Vragen die geschikt worden geacht voor opname in de huisartsgeneeskundige kennistoets, moeten voldoen aan bepaalde eisen. Ze moeten relevant zijn voor de praktijk en afgeleid zijn van huisartsgeneeskundige casuïstiek, ze dienen inhoudelijk correct te zijn en gebaseerd op adequate literatuur, en tenslotte moeten ze ondubbelzinnig zijn geformuleerd (Pollemans e.a., 1991; Pollemans e.a., in voorbereiding). De ervaring leert dat vrijwel alle vragen die voor gebruik in de toets worden aangeleverd, worden gewijzigd op grond van discussie in de beoordelingscommissie (Eekhof e.a., in voorbereiding). Deze ervaring bestaat ook elders (van Hessen & Verwijnen, 1987, 1989).

Over de tijd die het construeren van dergelijke toetsvragen kost, leert de ervaring het volgende. Na een introductie cursus van een dagdeel, en ondersteund door maandelijkse begeleidingsbijeenkomsten, bleken huisartsen die speciaal waren aangetrokken om dergelijke toetsvragen te construeren, gemiddeld 70 minuten nodig te hebben om één (casus)vraag te formuleren. Dit was inclusief het selecteren van een adequate literatuurverwijzing (Tan & Pollemans, 1992).

De *toetsafname* moet voldoen aan een aantal administratieve en organisatorische randvoorwaarden. De afname van een centraal vastgestelde en perifeer ingevulde toets moet tijdig worden geregeld en gepland. De deelnemers moeten tevoren zodanig zijn geïnstrueerd dat duidelijk is wat van hen wordt verwacht. De toets moet op tijd zijn gedrukt, en er moeten antwoordformulieren beschikbaar zijn. Het toetspakket, bestaande uit een toetsboekje, een voorgecodeerd antwoordformulier, invulinstructies en een retourenveloppe, moet op het juiste moment bij de deelnemers arriveren. Om anonieme verwerking van de antwoordformulieren te kunnen garanderen, moeten alle deelnemers een unieke code krijgen. Dat betekent dat administratieve bestanden moeten worden aangelegd van namen en adressen, en, daarvan gescheiden, een 'koppelbestand' met de codes.

Naast administratieve en organisatorische aspecten, geldt ten aanzien van de praktische uitvoerbaarheid van een algemene kennistoetsing bij huisartsen, ook de tijd die de huisartsen nodig hebben om de toets in te vullen, als een belangrijke factor. De invultijd die verschillende groepen deelnemers aangaven nodig te hebben, is weergegeven in tabel 4.3. Het betreft hier weer gegevens die zijn verzameld bij de kennistoetsafname van februari 1992.

Tabel 4.3. *Benodigde invultijd algemene huisartsgeneeskundige kennistoets door huisartsen (N=353).*

Deelnemers	Noodzakelijke invultijd				
	< 0.5 uur	0.5-1 uur	1-1.5 uur	1.5-2 uur	> 2 uur
huisartsen	–	9%	46%	37%	8%

De meeste huisartsen vulden de toets in ongeveer 1.5 tot 2 uur of minder in.

De derde optiek van waaruit de praktische uitvoerbaarheid van de algemene huisartsgeneeskundige kennistoetsing wordt benaderd, betreft de benodigde menskracht en middelen om de toetsgegevens te kunnen *analyseren en verwerken*, en *feedback* te kunnen samenstellen. Evenals bij de toetssamenstelling en -verzending vergt dit een aantal administratieve en organisatorische werkzaamheden. Randvoorwaarden zijn dat moet kunnen worden beschikt over adequate apparatuur en soft-ware om de gegevens geautomatiseerd te kunnen verwerken. Voor de verwerking van de gegevens van de huisartsen die deelnamen aan de kennistoets van februari 1992 is gebruik gemaakt van de middelen en procedures die beschikbaar waren bij het landelijk Samenwerkingsverband. Dat betekende dat de verwerkings- en analysetijd en de tijd die nodig is om individuele feedback te produceren, aanmerkelijk kon worden bekort.

In tabel 4.4 wordt samengevat welke activiteiten vanuit een centrale organisatie moeten worden uitgevoerd voor de productie van één landelijke huisartsgeneeskundige kennistoets. Tevens is een schatting gegeven van de tijdsinvestering die de verschillende activiteiten vergen. Daarbij wordt uitgegaan van een situatie waarin een gestandaardiseerde procedure bestaat voor de samenstelling van toetsen en de verwerking van toetsgegevens, waarin de noodzakelijke computerapparatuur beschikbaar is en de benodigde programmatuur niet meer hoeft te worden ontwikkeld.

Tabel 4.4. Activiteiten, menskracht en geschatte tijdsinvestering bij landelijke kennistoetsing.

Activiteiten	Uitvoering	Tijdsinvestering
Formuleren toetsvragen <ul style="list-style-type: none"> - formuleren toetsvragen - literatuurverwijzing completeren - bespreken relevantie, vorm en inhoud - toetsvragen in standaard-layout brengen 	inhoudsdeskundige inhoudsdeskundige inhoudsdeskundigen administratie	60* per vraag 10* per vraag 160 vragen: 2 dagdelen 160 vragen: 1 week
Toetsconstructie (toets van 160 vragen) <ul style="list-style-type: none"> - controle/aanvullen items per onderwerp blauwdruk - controle/aanvullen items nevenindelingen blauwdruk - redigeren eindversie toets - toets drukklaar maken - (reproductie toets) 	inhoudsdeskundige inhoudsdeskundige inhoudsdeskundige administratie (drukker)	2 dagdelen 2 dagdelen 2 dagdelen 2 dagdelen (PM)
Toetsafname <ul style="list-style-type: none"> - toetsafname plannen - deelnemers werven - administratief bestand van deelnemers aanleggen* - deelnemersinstructie formuleren* - voorcoderen antwoordformulieren en vragenlijsten** - individuele toetspakketten samenstellen en verzenden - antwoordsleutel samenstellen en verzenden 	inhoudsdeskundige administratie administratie inhoudsdeskundige administratie administratie administratie	PM PM per 100 pp: 1 dagdeel 1 dagdeel per 100 pp: 2 dagdelen per 100 pp: 2 dagdelen per 100 pp: 1 dagdeel
Analyse en verwerking toetsgegevens <ul style="list-style-type: none"> - antwoordformulieren invoeren via optisch lezer - foutencorrectie uitvoeren, verwijderen items - (item)analyse uitvoeren - invoeren opmerkingen deelnemers - samenstellen definitieve toets voor feedback - invoeren satisfactiegegevens deelnemers - samenstellen overzichten satisfactie 	administratie methodoloog methodoloog administratie inhoudsdeskundigen administratie administratie	1 dagdeel 1 dagdeel 1 dagdeel per 100 pp: 1 dagdeel 1 dagdeel per 100 pp: 1 dagdeel PM
Feedbackproductie en -verzending <ul style="list-style-type: none"> - productie: individuele toetsresultaten - samenstellen groepsoverzichten: gemiddelde scores - verzenden feedback 	adm./methodoloog adm./methodoloog administratie	2 dagdelen 1 dagdeel per 100 pp: 1 dagdeel
Coördinatie/management <ul style="list-style-type: none"> - procesbewaking 	inhoudsdeskundige	PM
Evaluatie/onderzoek <ul style="list-style-type: none"> - interpretatie gegevens en rapportage - gebruikersadviezen - onderzoek naar kwaliteitsverbetering instrument 	inhoudsdesk./method. inhoudsdesk./method. inhoudsdesk./method.	2 dagdelen PM PM
Diversen <ul style="list-style-type: none"> - bijvoorbeeld: informatie naar derden 	adm./inh.dsk./method.	PM

* eenmalige activiteit (bij herhaalde deelname aan toets)

** gebeurt bij reguliere toetsafname door deelnemers zelf

De tijdsinvestering voor de verschillende activiteiten in dit model is geschat. Een al te absolute betekenis kan er dan ook niet aan worden toegekend. Het model kan het beste worden gekenschetst als een managementmodel, waarbij is uitgegaan van een routine-toetsafname voor huisartsen-in-opleiding met extra deelnemende huisartsen. Het 'aanleggen administratief deelnemersbestand' is bij herhaalde deelneming slechts eenmaal nodig. De formulieren worden voor de routinematig deelnemenden niet centraal voorgecodeerd. Werkzaamheden in het kader van specifieke onderzoeksdoelen, bijvoorbeeld met betrekking tot kwaliteitsverbetering van de toets, zijn als PM-post opgenomen. Grondige evaluatie van het ontwikkelde instrument en van de praktische betekenis van de gegevens die ermee worden verzameld, is echter een voorwaarde om de kwaliteit van het instrument te kunnen waarborgen.

Zoals de tabel laat zien vergen de activiteiten die nodig zijn om een landelijke kennistoets te produceren en te verwerken de inzet van verschillende disciplines: administratie, methodologie en (huisartsgeneeskundige tezamen met toetstechnische) inhoudsdeskundigheid. De verschillende betrokkenen zijn, met piekperioden, actief in de perioden voorafgaand aan en na de feitelijke toetsafname.

De ervaring leert dat voor het construeren, afnemen en verwerken van één kennistoets van 160 vragen menskracht vanuit de verschillende disciplines gedurende ongeveer een half jaar moet worden ingezet. Huisartsgeneeskundig inhoudelijke deskundigheid is vooral van belang in de constructieperiode van de toets en op het moment van de definitieve vaststelling van de toets. Deskundigheid op het gebied van de methodologie is vooral nodig in de fase van de verwerking en analyse van de toetsgegevens, en bij het interpreteren van de betekenis die aan de resultaten mag worden gehecht. De overige werkzaamheden zijn voornamelijk van administratieve aard. Als er eenmaal een toets is geconstrueerd, is het aantal deelnemers alleen nog een factor die op het moment van de toetsafname en bij de samenstelling en verzending van de feedback, extra administratieve werkzaamheden met zich meebrengt. Het leeuwedeel van de tijd is dus nodig voor het constructieproces van de toetsvragen en het uitvoeren van standaard-administratieve handelingen ten behoeve van de centrale organisatie van de toetsafname. Het zal duidelijk zijn dat, als kan worden beschikt over een goed functionerende itembank die een efficiënt hergebruik van toetsvragen mogelijk maakt, in de constructietijd veel tijdwinst kan worden geboekt. Dit vooronderstelt, naast technische randvoorwaarden, echter ook de beschikbaarheid van een ruim toetsvragenbestand.

4.4. CONCLUSIES PRAKTISCHE TOEPASBAARHEID ALGEMENE HUISARTSGENEESKUNDIGE KENNISTOETSING

Geconcludeerd kan worden dat het goed mogelijk is een voor huisartsen relevante kennistoets samen te stellen en af te nemen, onder de conditie dat daarbij gebruik kan worden gemaakt van bestaande procedures voor wat betreft toetssamenstelling, gegevensverwerking en feedbackverzorging. Het construeren van goede toetsvragen is geen sinecure en vereist specifieke inhoudelijke en toetstechnische kennis en vaardigheden. De constructie van toetsvragen is bovendien een proces dat bij voorkeur wordt begeleid door een groep inhoudelijk deskundige vragenmakers.

De voor een landelijke toetsafname noodzakelijke administratieve en organisatorische procedures, maken de uitvoering van deze activiteiten door een landelijke bureau-organisatie die is voorzien van een adequate outillage, een vereiste.

Het invullen van een kennistoets van 160 vragen kost ervaren huisartsen gemiddeld ongeveer anderhalf uur. Gezien het relatief grote aantal huisartsen dat de toets te lang vond, lijkt deze toetslengte de maximaal haalbare. De huisartsen die aan een algemene kennistoets deelnamen, zijn in overgrote meerderheid enthousiast over de toets. Vragen en onderwerpen worden vrijwel unaniem als huisartsgeneeskundig relevant beoordeeld. De meeste huisartsen oordelen positief over de gebruikswaarde van de toets bij het onderkennen van kennislacunes of bij het bepalen van de keuze voor nascholings thema's.

DEEL II (hoofdstuk 5 en 6)

ONDERZOEK NAAR SAMENHANG KENNISNIVEAU EN ACHTERGRONDKENMERKEN HUISARTSEN

De tweede hoofddoelstelling van het onderzoek plus de bijbehorende vraagstellingen luiden als volgt.

Doelstelling II:

Bepaling van de samenhang van verschillende professionele kenmerken met de algemene kennis van huisartsen.

Vraagstellingen:

- II.1. Wat is het kennisniveau van verschillende groepen huisartsen op de algemene kennistoets als geheel en wat is het niveau op onderdelen van de kennistoets?
- II.2. Hoe hangt het kennisniveau van huisartsen samen met bepaalde persoonlijke of professionele achtergrondkenmerken van deze huisartsen?
- II.3. In hoeverre hangt het oordeel dat huisartsen zelf aangeven over hun specifieke expertise samen met hun kennisniveau?

In de volgende hoofdstukken wordt beschreven op welke wijze en met welk resultaat de vragen behorend bij dit deel van het onderzoek zijn beantwoord. De gegevens zijn verzameld door een op de blauwdruk afgestemde kennistoets voor te leggen aan groepen praktiserende huisartsen. (Zie voor een beschrijving van de kennistoets hoofdstuk 2.) Tevens is hen gevraagd een vragenlijst naar achtergrondkenmerken in te vullen. In hoofdstuk 5 wordt beschreven wat de bevindingen waren met betrekking tot de samenhang tussen kennisniveau en achtergrondkenmerken. In hoofdstuk 6 komt het verband aan de orde tussen de kennis van huisartsen en hun eigen inschatting van hun expertise.

Hoofdstuk 5

ALGEMEEN KENNISNIVEAU IN RELATIE TOT ACHTERGRONDKENMERKEN VAN ERVAREN HUISARTSEN

SAMENVATTING

Het eventueel verband tussen de algemene kennis van huisartsen en hun achtergrondkenmerken is onderzocht door de gemiddelde scores van ervaren huisartsen op een algemene kennistoets, te relateren aan persoonskenmerken, demografische kenmerken en professionele kenmerken van deze huisartsen.

De gemiddelde score van ervaren huisartsen ligt ongeveer op het niveau van aanstaande huisartsen na driekwart van hun tweejarige opleiding. Naarmate huisartsen ouder zijn, is hun gemiddelde kennistoetsscore lager. Lidmaatschap van het NHG en opleiderservaring vertonen een significant positief verband met de gemiddelde kennistoetsscore. In tegenstelling tot de verwachtingen is geen verband aangetroffen tussen werkverband of praktijksituering en kennisniveau.

Een stapsgewijze multipele regressie-analyse werd uitgevoerd om te kunnen nagaan welke variabelen het meeste bijdragen aan de verklaarde variantie van de toetsscore van de huisartsen. Het totale percentage verklaarde variantie bedroeg 13. Het overgrote deel van de variantie in kennistoetsscore kan dus niet worden verklaard uit verschillen in achtergrondkenmerken van de huisartsen.

Het verloop van de scores voor met name medisch-inhoudelijke thema's doet vermoeden dat de kennis daarvan nauw samenhangt met de mate waarin huisartsen er in de praktijk mee worden geconfronteerd. Dit geldt met name voor aandoeningen van het bewegingsapparaat of huidaandoeningen. Hetzelfde patroon wordt gezien voor de vragen over 'medicamenteus beleid'. Het vermoeden dat kennis over bepaalde onderwerpen en daadwerkelijke praktijkervaring met deze thema's samenhangen, wordt ondersteund door de bevinding dat huisartsen met praktijken met relatief veel ouderen, gemiddeld significant hogere scores behaalden op de vragen die betrekking hebben op 'ouderen (boven 75 jaar)' en 'chronische aandoeningen'. Huisartsen met veel jonge gezinnen in de praktijk scoren gemiddeld wel hoger dan hun collega's op de categorie 'jongeren beneden 15 jaar', maar dit verschil was niet significant.

5.1. INLEIDING

Om eventuele verbanden tussen achtergrondkenmerken van huisartsen en hun kennisniveau te achterhalen, werd een grotendeels explorierend, deels hypothese-toetsend onderzoek uitgevoerd. Hiervoor werden de gegevens gebruikt die zijn verzameld in dezelfde toetsafname als beschreven in Deel I in de paragrafen over constructvalidering, betrouwbaarheid en acceptabiliteit van de algemene huisartsgeneeskundige kennistoets. Deze toets was samengesteld volgens de blauwdruk. De deelnemende huisartsen beantwoordden tevens een vragenlijst over achtergrondkenmerken en een satisfactie-vragenlijst. De vragenlijst over achtergrondkenmerken bevatte alle kenmerken waarover de literatuur aanwijzingen gaf dat ze zouden kunnen samenhangen met de kennis van huisartsen. De satisfactie-vragenlijst diende vooral om de acceptabiliteit van de toets voor huisartsen na te gaan. Van enkele variabelen uit deze vragenlijst werd tevens nagegaan of ze samenhang vertoonden met de kennistoetsscores van de huisartsen.

In dit hoofdstuk wordt het resultaat beschreven.

Onderzoeksvraag II.1.:

Wat is het kennisniveau van huisartsen op de algemene kennistoets als geheel en wat is het niveau op onderdelen van de kennistoets?

Onderzoeksvraag II.2.:

Hoe hangt het kennisniveau van huisartsen samen met bepaalde persoonlijke of professionele achtergrondkenmerken van deze huisartsen?

Opzet onderzoek:

- bepaling variatie in kennis bij ervaren huisartsen, huisartsen-in-opleiding en co-assistenten huisartsgeneeskunde, uitgedrukt in goed-min-foutscore op totale algemene kennistoets en in score op onderdelen toets
- bepaling samenhang algemene kennis en achtergrondvariabelen met behulp van correlatie- en variantieanalyses

5.2. METHODE ONDERZOEK SAMENHANG KENNIS EN ACHTERGRONDKENMERKEN

5.2.1. Opzet onderzoek

Om een indruk te krijgen van het algemene kennisniveau van huisartsen, werd een algemene huisartsgeneeskundige kennistoets afgenomen bij huisartsen. De toetsafname vond plaats tezamen met de reguliere toetsafname bij huisartsen-in-opleiding in februari 1992. Aan de toetsafname namen ook medische studenten (co-assistenten huisartsgeneeskunde) deel. Dat betekende dat de scores konden worden vergeleken van groepen in diverse fasen van opleiding en ervaring op het gebied van de huisartsgeneeskunde. De gemiddelde toetsscores plus standaarddeviaties van alle deelnemers werden, conform de gebruikelijke procedures, uitgedrukt in percentages 'goed-min-fout'. Toetsscores zijn berekend over de totale kennistoets en op deelgebieden van de toets, en worden in beschrijvende statistieken weergegeven.

De vraag die in dit onderzoeksdeel wordt beantwoord, is of kenmerken van bepaalde groepen huisartsen samengaan met een afwijkend kennisniveau. Dit kan zowel het algemene

kennisniveau, uitgedrukt als de score op de totale toets, betreffen als het kennisniveau over bepaalde onderdelen van de toets. Gezocht wordt naar (combinaties van) achtergrondvariabelen die relatief meer of minder bijdragen aan deze verschillen. De *afhankelijke variabelen* in dit deel van het onderzoek vormden de gemiddelde goed-min-foutscores van (sub)groepen van huisartsen op de totale kennistoets of op onderdelen daarvan. De *onafhankelijke variabelen* vormen verschillende achtergrondkenmerken van de huisartsen en satisfactievariabelen.

5.2.2. Onderzoeksinstrumenten en variabelen

- *algemene kennistoets*

De verdeling van de 160 juist-onjuist-vraagtekenvragen in de kennistoets was geheel conform de hoofdingeling van de blauwdruk (zie Bijlage 5 voor een voorbeeldpagina van de algemene huisartsgeneeskundige kennistoets van februari 1992). De nevenindelingen volgden grotendeels de streefgetallen van de blauwdruk. De toets bevatte ongeveer 10 vragen meer over medicamenteus beleid, en ongeveer 10 vragen minder over diagnostiek dan de streefgetallen van de blauwdruk aangaven.

De huisartsen vulden de antwoorden op de toetsvragen in op optisch leesbare formulieren. Deze waren om privacy-overwegingen voorgedrukt. De deelnemers aan de toets werd tevens gevraagd om commentaar te geven op de toetsvragen. Op basis hiervan werd de definitieve toets, waarover de nadere analyses zijn uitgevoerd en waarover feedback werd gegeven, vastgesteld op 146 vragen. Dat betekent dat 14 vragen uit de toets zijn verwijderd. De deelnemende huisartsen ontvingen ongeveer zes weken nadat ze de toetsformulieren hadden ingestuurd, individueel feedback over hun toetsresultaten. Deze bestond uit een overzicht van de eigen toetsresultaten: percentage goede, foute en met vraagteken beantwoorde vragen over de complete toets en over de onderdelen waaruit de toets bestond. Deze werden vergezeld van de gemiddelde resultaten van de eigen (opleiders-, nascholings-, toetsings)groep en de gemiddelde resultaten van alle deelnemende huisartsen.

- *achtergrondkenmerken-vragenlijst*

De persoonlijke, demografische en professionele variabelen werden verwerkt in een 'vragenlijst achtergrondkenmerken'. Deze vragenlijst bestond uit 19 vragen, zoveel mogelijk in een gesloten vorm (zie Bijlage 6). Ten behoeve van een anonieme verwerking waren alle vragenlijsten voorgedrukt. De belangrijkste variabelen worden hierna kort toegelicht.

persoonsvariabelen: leeftijd, geslacht.

Persoonskenmerken werden opgenomen om de aan de kennistoetsing deelnemende huisartsen te kunnen beschrijven. De literatuur geeft aan dat lagere scores mogen worden verwacht met toenemende leeftijd (Day et al, 1988; Steel et al, 1989; Norcini & Dawson-Saunders, 1994).

demografische praktijkvariabelen: werkverband, praktijksituering, praktijkgrootte, aantal patiëntcontacten, bijzondere kenmerken patiëntenpopulatie.

Demografische praktijkvariabelen worden in de literatuur in verband gebracht met het kennisniveau van artsen (zie hoofdstuk 1). Om de vraag te kunnen beantwoorden of bijzondere kenmerken van de praktijkpopulatie of speciale belangstelling voor bepaalde thema's bij huisartsen het kennisniveau beïnvloeden, zijn ook hierover vragen opgenomen.

professionele variabelen: kenmerken huisartsopleiding (plaats en duur), beroepservaring (duur werkervaring als huisarts, werktijdromvang), lidmaatschap beroepsvereniging, onderwijservaring (soort en duur), betrokkenheid in nascholing (tijdsbesteding cursorische nascholing en lezen vakliteratuur, lidmaatschap toetsingsgroep).

Niet alle huisartsen hebben de huisartsopleiding gevolgd, die immers pas sinds 1974 verplicht werd ingevoerd. Een eventuele samenhang van het al dan niet gevolgd hebben van de huisartsopleiding hangt uiteraard direct samen met leeftijd, maar toch leek het interessant om na te gaan of, apart van een eventueel leeftijdeffect, er verband kon worden gevonden. Ditzelfde gold voor praktijkervaring. Denkbaar was daarnaast een relatie tussen werkbelasting en kennisniveau: naarmate de werkbelasting minder is zou er meer tijd kunnen overblijven voor nascholing (Bloemer & Poiesz, 1987).

Lidmaatschap van met name een wetenschappelijke beroepsvereniging, zou kunnen wijzen op gerichte belangstelling voor ontwikkelingen in het vakgebied en wellicht samenhang vertonen met een relatief hoog kennisniveau. De literatuur geeft hiervoor aanwijzingen (Davis et al, 1990). Lidmaatschap van een toetsingsgroep zou, om dezelfde reden, kunnen samenhangen met meer kennis. Uit de literatuur kan worden afgeleid dat tijd besteed aan onderwijs en onderzoek, samenhang vertoont met het kennisniveau van artsen (Steel et al, 1989). Het leek daarom waarschijnlijk dat ervaring met het opleiden van (huis)artsen danwel ervaring als onderwijsgever in de nascholing, samenhang zou vertonen met het kennisniveau. Volgens Forrest et al (1989) beïnvloedt het opleiderschap in elk geval de houding tot nascholing positief. Van de hoeveelheid tijd die aan nascholing wordt besteed, met name de tijd die wordt besteed aan het bijhouden van de vakliteratuur, werd op grond van de literatuur verwacht dat deze samenhangt met het kennisniveau (Bertram & Bertram-Brooks, 1977).

- satisfactie-vragenlijst

De satisfactie-variabelen werden verwerkt in een vragenlijst bestaande uit 7 vijfpuntsstellingen (te beantwoorden met geheel eens, eens, neutraal, oneens, of geheel oneens) en enkele open vragen (zie Bijlage 7). De gegevens uit deze vragenlijst werden ook gebruikt bij de beantwoording van de vragen naar de acceptabiliteit van de algemene kennistoets (zie hoofdstuk 4). Ook deze vragenlijst werd ten behoeve van anonieme verwerking van de gegevens voorgedcodeerd.

De open vragen boden de gelegenheid commentaar te geven op de toets als geheel en op individuele toetsvragen.

5.2.3. Onderzoeksgroep

Omdat de onderzoeksvragen er niet op waren gericht uitspraken te doen over dé kennis van dé Nederlandse huisarts, is geen poging gedaan tot systematische werving van een representatieve steekproef uit de Nederlandse populatie van huisartsen. De representativiteit van de deelnemersgroep kan wel een rol spelen in de discussie over de vraag naar de toepasbaarheid van de algemene kennistoets in de nascholing van huisartsen. Om een eventuele selectiebias na te gaan, is een aantal achtergrondkenmerken van de deelnemers vergeleken met bekende gegevens van de landelijke populatie van huisartsen. Voor de landelijke gegevens werden de cijfers uit CBS- en NIVEL-registraties gebruikt (Hingstman & Pool, 1992).

De algemene kennistoets werd in de eerste helft van februari 1992 ingevuld door 365 ervaren huisartsen, alsmede door 108 studenten tijdens het co-assistentenschap huisartsgeneeskunde, en 446 huisartsen-in-opleiding. Deelnemende huisartsen waren afkomstig uit het gehele land. In verband met de noodzakelijke deelname van huisartsopleiders voor het parallelonderzoek naar de invloed van opleidingskenmerken op de kennisontwikkeling van huisartsen-in-opleiding, bestond de onderzoekspopulatie uit relatief veel huisartsopleiders. De opleiders zijn geworven via de afdelingen huisartsopleiding van de huisartsinstituten. Gevestigde huisartsen buiten het circuit van huisartsopleiders zijn geworven door inschakeling en tussenkomst van contactpersonen (stafleden LHV, inhoudscoördinatoren deskundigheidsbevordering, hoogleraren huisartsgeneeskunde). Dat leidde tot deelname van een zevental huisartsen-(waarneem)groepen. Daarnaast zegde een achttal toetsingsgroepen deelname toe. De huisartsen-in-opleiding waren verdeeld over alle zes opleidingsfasen, en afkomstig van alle acht huisartsopleidingen. De co-assistenten waren afkomstig van zes van de acht medische faculteiten. Tabel 5.1 geeft een overzicht van de deelnemers aan de algemene huisartsgeneeskundige kennistoets in februari 1992. Vanwege het verschil in wervingsprocedure worden opgave en deelname van opleiders en niet-opleiders apart vermeld.

Tabel 5.1. Deelnemers algemene kennistoets van februari 1992.

Deelnemers	Opgave aantal	Deelname aantal	Respons
<i>Huisartsen</i>			
opleiders	308	232	75%
niet-opleiders	167	133	80%
Totaal	475	365	77%
<i>Huisartsen-in-opleiding*</i>			
fase 1: 1 maand in opleiding		85	
fase 2: 5 maanden in opleiding		103	
fase 3: 9 maanden in opleiding		46	
fase 4: 13 maanden in opleiding		55	
fase 5: 17 maanden in opleiding		99	
fase 6: 21 maanden in opleiding		57	
Totaal		446	
<i>Co-assistenten huisartsgeneeskunde*</i>			
Totaal		108	

* alleen deelnamecijfers omdat geen individuele opgave plaatsvond. Groepen werden door groepsbegeleiders voor deelname aangemeld.

Uit tabel 5.1 is af te lezen dat uiteindelijk ingevulde antwoordformulieren binnenkwamen van 365 van de 475 ervaren huisartsen die waren opgegeven voor deelname. Dat betekent een respons van 77%.

Alle co-assistenten en huisartsen-in-opleiding, en ruim de helft van de ervaren huisartsen (208 huisartsen) vulden de toets in onder examenomstandigheden op één van de afdelingen huisartsopleiding, dat wil zeggen onder toezicht en zonder onderling overleg. De deelnemers

die de toets thuis invulden (157 huisartsen), werd dringend verzocht zich aan examencondities te houden, dat wil zeggen niet te overleggen en geen gebruik te maken van naslagwerken (zie Bijlage 8 voor de brief waarmee de deelnemende huisartsen werden geïnformeerd over de procedure bij de kennistoetsing). Een eventueel effect van de omstandigheden waaronder de toets is gemaakt op de betrouwbaarheid, is nagegaan door de resultaten van de onder toezicht deelnemende huisartsen te vergelijken met de resultaten van de 'thuiswerkers'. De betrouwbaarheidscoëfficiënt en de gemiddelde scores voor beide groepen bleken ongeveer gelijk. De conclusie was dat de meetfout voor beide groepen ongeveer hetzelfde was, zodat de resultaten van de groepen die de toets onder examencondities invulden niet anders gekarakteriseerd hoefden te worden dan de resultaten van de groepen die de toets thuis invulden (Düsmen, 1992).

De huisartsen ontvingen de feedback op de toetsresultaten ongeveer zes weken nadat ze hun formulieren hadden ingestuurd (zie Bijlage 9 voor een geanonimiseerd voorbeeld plus toelichtende brief).

Alle huisartsen vulden ook de achtergrondvragenlijst in. Twaalf huisartsen gaven op die lijst aan ofwel geen patiëntcontacten te hebben en/of niet als huisarts werkzaam te zijn. Deze huisartsen werden uit het bestand gehaald. Het onderzoek naar samenhangen tussen achtergrondkenmerken en kennis van huisartsen is gebaseerd op de resterende 353 huisartsen.

Tabel 5.2. Achtergrondkenmerken onderzoeksgroep ervaren huisartsen, die deelnamen aan algemene kennistoets februari 1992, plus landelijke cijfers.

Achtergrondkenmerken	Deelnemers algemene kennistoets (N=353) (%)	Alle Nederlandse huisartsen (N=6535) (NIVEL, 1992) (%)
<i>Geslacht</i>		
man	88	86
vrouw	12	14
<i>Leeftijdsverdeling</i>		
< 34 jaar	5	8
35-39 jaar	25	25
40-44 jaar	35	28
45-49 jaar	22	18
50-54 jaar	8	8
55-59 jaar	4	6
> 60 jaar	1	6
<i>Praktijksituering</i>		
grote stad (> 100.000)	25	25
stad	27	27
verstedelijk platteland	31	37
platteland	17	11
<i>Werkverband</i>		
solist	45	53
duo/associatie	29	31
groepspraktijk/gezondheidscentrum	26	16

Tabel 5.2 geeft een beeld van enkele persoonlijke en demografische achtergrondkenmerken van de onderzoeksgroep. Ter vergelijking zijn landelijke cijfers opgenomen.

Hoewel niet is gestreefd naar een representatieve afspiegeling van de Nederlandse huisartsen, geven de achtergrondkenmerken van de huisartsen geen aanleiding te veronderstellen dat de groep deelnemers belangrijk afwijkend is van de doorsnee-populatie huisartsen in Nederland. Wat betreft leeftijd is de 'midden'groep (tussen 40 en 50 jaar) enigszins oververtegenwoordigd. Dit wordt waarschijnlijk veroorzaakt door het relatief grote aantal opleiders dat de totale groep telt. Opleiders zijn waarschijnlijk gemiddeld ouder dan de overige huisartsen, omdat ze minstens vijf jaar praktijkervaring moeten hebben voordat ze als opleider werkzaam kunnen zijn. De onderzoeksgroep telt naar verhouding veel huisartsen die werkzaam zijn op het platteland. Tabel 5.2 laat ook zien dat er gemiddeld minder solisten en meer huisartsen in groepspraktijken of gezondheidscentra onder de deelnemers aan de kennistoets zijn, dan in de totale populatie van Nederlandse huisartsen. Wellicht wordt ook dit veroorzaakt door het grote aantal opleiders onder de deelnemers.

5.2.4. Analyses

De afhankelijke variabele, de gemiddelde kennistoetsscore, is uitgedrukt als percentage van de maximaal te behalen goed-min-foutscore. Om samenhangen tussen de onafhankelijke variabelen en de kennisscores te achterhalen, zijn frequentietabellen samengesteld van de belangrijkste onafhankelijke variabelen. Vervolgens zijn bivariate samenhangen van de verschillende achtergrondvariabelen met de kennisscore bepaald op basis van variantie-analyses. Daarbij werd nagegaan of de verschillen in gemiddelde score (de afhankelijke variabele) voor de categorieën van de op een nominale schaal gemeten onafhankelijke variabelen, statistisch significant waren (van Eijk & Gubbels, 1987). Pearson's produkt-moment correlatiecoëfficiënten werden berekend om de samenhang na te gaan tussen de kennisscore en de variabelen gemeten op interval- of rationiveau (Meerling, 1981; Slotboom, 1987). Zoals gebruikelijk is bij deze gesloten, schriftelijke, 'objectieve' toetsvorm, is er daarbij van uitgegaan dat de afhankelijke variabele, de kennistoetsscore, werd gemeten op een intervalschaal (Foulkes, 1994). De satisfactiescores werden berekend als gemiddelde percentages voor de antwoordcategorieën per satisfactievariabele. Voor de samenhang van de satisfactiescores met de kennisscore werden eveneens Pearson produkt moment correlatiecoëfficiënten berekend.

Om de vraag te kunnen beantwoorden welke variabelen de meeste voorspellende waarde hebben ten aanzien van de kennistoetsscore, werd een stapsgewijze multiële regressie-analyse uitgevoerd. Daartoe werd een selectie gemaakt van de onafhankelijke variabelen, waarvan op theoretische gronden werd verwacht, dat ze het meest zouden bijdragen aan de kennistoetsscores.

De volgende variabelen werden, zoveel mogelijk gedichotomiseerd, in de regressievergelijking opgenomen:

- leeftijd,
- aantal jaren werkervaring,
- gevolgd hebben huisartsopleiding (al dan niet),
- lidmaatschap wetenschappelijke beroepsvereniging (al dan niet NHG-lid),
- lidmaatschap toetsingsgroep (al dan niet).
- omvang nascholing per jaar,
- ervaring als opleider (al dan niet),

- werkverband (al dan niet solist),
- geografische situering praktijk,
- specifieke kenmerken praktijkpopulatie (doorsnee of afwijkend),
- aantal patiënten in praktijk,
- aantal patiëntcontacten per week.

5.3. RESULTATEN SAMENHANG ALGEMENE HUISARTSGENEESKUNDIGE KENNIS EN ACHTERGRONDKENMERKEN

5.3.1. Gemiddelde kennistoetsscores

Tabel 5.3. Gemiddelde goed-min-foutscores van ervaren huisartsen op (hoofdstukken uit) de algemene kennistoets.

Onderwerpen	Aantal vragen	Gemiddelde goed-min-foutscore (%)							
		co's	huisartsen-in-opleiding	huisartsen*					
				1	2	3	4	5	totaal
		N 108	N 445	N 22	N 68	N 99	N 109	N 53	N 351
Algemeen	7	17	35	48	58	45	45	40	47
Bloed	2	63	74	73	67	62	66	67	66
Tr. digestivus	9	16	48	50	53	50	46	37	48
Oog	6	14	26	33	36	26	19	20	26
Oor	5	46	54	52	57	55	45	44	50
Tr. circulatorius	14	18	46	48	54	55	49	44	51
Bewegingsapparaat	18	13	30	34	39	37	38	34	37
Neurologie	8	14	24	34	29	22	22	14	23
Psychische problemen	7	21	38	29	34	35	35	31	34
Tr. respiratorius	15	31	52	63	61	62	64	56	62
Huid	10	32	48	57	57	62	55	64	59
Endocr/metab/voeding	6	34	55	64	61	57	53	66	58
Urologie	6	8	40	41	40	39	37	32	38
Zwang/bev/anticonc.	5	-5	23	39	33	29	29	32	31
Vr. geslachtsorganen	6	14	15	15	21	21	12	10	16
Mn. geslachtsorganen	5	19	44	49	57	36	42	42	44
Theoretische onderw.	17	40	51	59	54	55	48	50	52
Totaal	146	23	41	47	48	46	43	41	45
std		9	11	8	13	10	12	12	11

* huisartsen ingedeeld naar werkervaring: 1 = < 5 jr; 2 = 5-10 jr; 3 = 10-15 jr; 4 = 15-20 jr; 5 = > 20 jr.

In tabel 5.3 wordt een overzicht gegeven van de gemiddelde scores van de huisartsen per toetsonderwerp. De huisartsen zijn in groepen ingedeeld naar aantal jaren werkervaring. Van twee huisartsen was de omvang van de werkervaring niet bekend. De gegevens zijn daarom

gebaseerd op de resterende 351 huisartsen. Ter vergelijking zijn de gemiddelde scores van de co-assistenten en huisartsen-in-opleiding opgenomen.

De gemiddelde goed-min-foutscore van de huisartsen op de algemene kennistoets bedroeg 45%, met een standaarddeviatie van 11%. Deze score is gebaseerd op de geschoonde toetsversie van 146 vragen. De hoogste goed-min-foutscore die door een ervaren huisarts werd behaald bedroeg 76%, de laagste score bedroeg 15%.

Voor de interpretatie van de gegevens uit tabel 5.3 moet rekening worden gehouden met het volgende. In de eerste plaats betreft het hier gegevens uit cross-sectioneel onderzoek: verschillende groepen worden op één moment in de tijd met elkaar vergeleken. Verder moet er rekening mee worden gehouden dat de precisie van de gegevens niet voor elke subtoets (onderwerp), resp. subgroep van deelnemers even groot is, maar afhankelijk van de omvang van elk van beide.

De scoreverschillen tussen de diverse groepen deelnemers aan de toets zijn op het niveau van de totale toets significant ($p < 0.05$). De gemiddelde score van 45% van de ervaren huisartsen komt overeen met het niveau van huisartsen-in-opleiding na anderhalf jaar van de tweejarige opleiding (van Leeuwen, in voorbereiding). Met toenemende praktijkervaring na afloop van de opleiding dalen de gemiddelde totaalscores op de kennistoets. Vergelijking van het scorepatroon op het niveau van de separate onderwerpen tussen co-assistenten, huisartsen-in-opleiding en ervaren huisartsen (verdeeld in groepen met een toenemend aantal jaren ervaring) laat enkele opvallende verschillen zien. Voor sommige onderwerpen geldt dat ook de co-assistenten al relatief hoge scores bereiken (bijvoorbeeld voor het onderwerp 'huisartsgeneeskundig-theoretische aspecten'). Gedurende de huisartsopleiding en bij de ervaren huisartsen blijft de kennis op dit gebied ongeveer op hetzelfde niveau. De ervaren huisartsen bereiken, in vergelijking met huisartsen-in-opleiding, relatief hoge scores op het onderwerp 'algemeen' (hieronder vallen bijvoorbeeld zaken als 'bijwerkingen van geneesmiddelen', 'koorts met onbekende oorzaak'), 'tractus respiratorius', 'tractus circulatorius', 'bewegingsapparaat', 'huid' en 'zwangerschap/bevalling en anticonceptie'. Op andere onderwerpen, zoals 'bloed' (overigens met slechts 2 vragen in de toets vertegenwoordigd), 'oor' of 'psychische problemen' zijn de gemiddelde scores van de ervaren huisartsen lager dan die van de huisartsen-in-opleiding.

In tabel 5.4 worden de totaalscores en de deelscores voor de verschillende groepen deelnemers weergegeven voor de onderdelen in de toets, die corresponderen met de nevenindelingen van de blauwdruk. Voor de significantie van de scoreverschillen tussen de verschillende groepen, geldt hetzelfde als bij de bespreking van de vorige tabel, zij het dat de omvang van de verschillende (sub)categorieën van de nevenindelingen in het algemeen groter is dan de categorieën uit de hoofdingeling.

Uit tabel 5.4 blijkt dat ook voor de onderwerpen, die de nevenindelingen van de blauwdruk representeren, geldt dat er in het algemeen een stijgende lijn te zien is in de kennisscores tussen de co-assistenten, huisartsen-in-opleiding en ervaren huisartsen. De hoogste gemiddelde scores worden aangetroffen bij de meest recent afgestudeerde huisartsen.

Tabel 5.4. Gemiddelde goed-min-foutscores van ervaren huisartsen op onderdelen van de algemene kennistoets, corresponderend met nevenindelingen uit de blauwdruk.

Onderwerpen nevenindelingen kennistoets	Aantal vragen	Gemiddelde goed-min-foutscore (%)							
		co's	huis-artsen-in-opleiding	huisartsen*					
				1	2	3	4	5	totaal
				N	N	N	N	N	N
		108	445	22	68	99	109	53	351
<i>Leeftijdscategorieën patiënten</i>									
< 15 jaar	35	35	54	58	62	59	58	55	59
15-75 jaar	78	14	33	40	40	38	35	33	37
> 75 jaar	16	36	40	42	50	47	45	40	46
niet leeftijdgebonden	17	40	51	59	54	55	48	50	52
<i>Aspecten consult</i>									
diagnostiek	48	27	41	47	47	43	41	40	43
medicamenteus beleid	43	9	35	40	46	45	42	38	43
niet medicamenteus beleid	31	21	43	46	48	44	43	39	44
epidemiologie	9	43	51	57	59	54	49	51	53
rest	15	39	51	58	55	56	48	50	53
<i>Spoedgevallen</i>									
spoed	12	26	45	55	55	48	43	43	48
niet-spoed	134	23	41	46	48	45	43	41	44
<i>Chronische aandoeningen</i>									
chronisch	30	15	39	46	47	47	45	46	46
niet-chronisch	116	25	42	47	49	46	43	40	45
Totaal std	146	23	41	47	48	46	43	41	45
		9	11	8	13	10	12	12	11

* huisartsen ingedeeld naar werkervaring: 1 = < 5 jr; 2 = 5-10 jr; 3 = 10-15 jr; 4 = 15-20 jr; 5 = > 20 jr.

De co-assistenten bereiken, in vergelijking met hun eigen totale gemiddelde score en in vergelijking met de gemiddelde scores van huisartsen-in-opleiding en ervaren huisartsen, gemiddeld relatief hoge scores op het gebied van de diagnostiek en gemiddeld relatief lage scores voor medicamenteus beleid. Wat betreft diagnostiek en beleid lopen de gemiddelde scores van de ervaren huisartsen voor deze categorieën weinig uit elkaar. Dit in tegenstelling tot de gemiddelde scores van co-assistenten en huisartsen-in-opleiding voor deze categorieën. Op de vragen over patiënten jonger dan 15 jaar bereiken alle deelnemende groepen relatief hoge scores. Op de vragen die betrekking hebben op patiënten in de leeftijdscategorie tussen 15 en 75 jaar, behalen alle deelnemende groepen relatief lage scores in vergelijking met hun eigen groepsgemiddelden. De trend in de scores is in deze subcategorieën van toetsvragen, voor de onderscheiden groepen wel gelijk aan de algemene trend.

Wat opvalt is dat de gemiddelde score voor de groep vragen die 'spoedgevallen' betrof, het hoogst is voor de huisartsen die in de eerste 10 jaar van hun beroeps carrière verkeren. De gemiddelde score voor chronische aandoeningen blijft, in tegenstelling tot het patroon op de totale toets, bij huisartsen met toenemende ervaring op hetzelfde niveau.

5.3.2. Samenhang achtergrondkenmerken en kennisniveau

In tabel 5.5 wordt het resultaat weergegeven van de variantie-analyses die zijn uitgevoerd met de achtergrondvariabelen die op een nominale schaal zijn gemeten, om een eventueel verband met de kennistoetsscore te achterhalen. De F-toets bij de variantie-analyse drukt uit of de gemiddelde goed-min-foutscores van de onderscheiden categorieën significant van elkaar verschillen.

Tabel 5.5. Samenhang van verschillende categorieën van achtergrondvariabelen van huisartsen met het kennisniveau, uitgedrukt in de gemiddelde goed-min-foutscore (plus standaarddeviatie) op de algemene kennistoets.

Achtergrondvariabelen	Aantal	Gemiddelde goed-min-fout-score (%)	Standaarddeviatie (%)	Resultaat variantie-analyse
<i>Geslacht</i>				
man	312	45	11	Pr>F: 0.7760 R ² : 0.000
vrouw	41	45	10	
<i>Huisartsopleiding</i>				
wel	205	47	11	Pr>F: 0.0001* R ² : 0.077
niet	148	42	12	
<i>Plaats huisartsopleiding</i>				
Amsterdam: UvA	23	49	8	Pr>F: 0.3717 R ² : 0.037
Amsterdam: VU	25	49	10	
Groningen	20	45	13	
Leiden	19	49	10	
Maastricht	24	46	11	
Nijmegen	46	44	12	
Rotterdam	14	44	11	
Utrecht	33	45	11	
<i>Lidmaatschap beroepsvereniging</i>				
NHG-leden	235	46	11	Pr>F: 0.0001* R ² : 0.058
niet NHG-leden	116	41	12	
<i>Opleiderservaring</i>				
wel	280	45	11	Pr>F: 0.0306* R ² : 0.013
niet	71	42	13	

Tabel 5.5. vervolg

<i>Type opleiderservaring</i>				
alleen artsopleiding	25	49	11	Pr>F: 0.0548 R ² : 0.039
alleen huisartsopleiding	128	44	12	
alleen nascholing	21	47	12	
arts- + huisartsopleiding	36	43	8	
artsopleiding + nascholing	13	43	10	
huisartsopleiding + nascholing	38	48	10	
arts-/huisartsopleiding + nascholing	19	49	9	
<i>Lidmaatschap toetsingsgroep</i>				
wel	179	46	11	Pr>F: 0.1111 R ² : 0.007
niet	172	44	12	
<i>Werkverband</i>				
solo	158	45	12	Pr>F: 0.0836 R ² : 0.019
duo/associatie	102	43	11	
groepspraktijk	39	46	12	
gezondheidscentrum	51	48	9	
<i>Praktijksituering</i>				
grote stad	86	45	12	Pr>F: 0.9280 R ² : 0.001
stad	97	44	11	
verstedelijkt platteland	109	45	12	
platteland	60	46	11	
<i>Bijzondere kenmerken praktijk</i>				
doorsnee-praktijk	186	44	12	Pr>F: 0.5113 R ² : 0.007
veel ouderen	61	47	11	
veel jonge gezinnen	52	45	11	
overig bijzonder	54	44	11	

* significant verschil

Voor de variabelen 'huisartsopleiding', 'lidmaatschap beroepsvereniging' en 'opleiderservaring' werd een significant verband met de kennistoetsscore gevonden ($p < 0.05$). Geen van de overige variabelen leverde een significant verband op met de kennistoetsscore. Voor een aantal variabelen waren de verschillen in gemiddelden tussen de categorieën waaruit de variabelen bestonden, overigens wel in de richting die werd verwacht: huisartsen die lid zijn van een toetsingsgroep behalen bijvoorbeeld gemiddeld iets hogere scores dan de huisartsen die dat niet zijn. Tegen de verwachting in werd geen relatie gevonden tussen een (solistisch) werkverband en de kennistoetsscore. Huisartsen in duo- of associatiepraktijken behaalden gemiddeld zelfs een iets lagere score dan de solisten.

De samenhang tussen het gemiddelde kennisniveau van huisartsen met bijzondere praktijkkenmerken en de resultaten op de 'subtoetsen' met betrekking tot 'chronische aandoeningen' en 'leeftijdscategorieën van patiënten' is nader onderzocht. Daarbij is gezocht naar een eventueel verband tussen bijvoorbeeld kennis over jongeren bij huisartsen met relatief veel jonge gezinnen in de praktijk, of naar verband met kennis over ouderen bij

huisartsen met relatief veel ouderen in de praktijk. De gemiddelden worden in tabel 5.6 weergegeven.

Tabel 5.6. *Samenhang gemiddelde goed-min-foutscore van huisartsen op onderdelen kennistoets en enkele praktijkkenmerken van huisartsen.*

Onderdelen kennistoets	Gemiddelde goed-min-foutscore van groepen huisartsen met speciale praktijkkenmerken (in percentages, standaarddeviatie tussen haakjes)			
	doorsneepraktijk (N=186)	veel ouderen (N=61)	veel jonge gezinnen (N=52)	overig bijzonder (n=54)
<i>Chronische aandoeningen</i> (N=30 vragen)	44 (18)	47 (18)	42 (19)	39 (19)
<i>Leeftijdscategorieën patiënten</i>				
< 15 jaar (N=35 vragen)	59 (14)	59 (14)	61 (15)	55 (14)
15-75 jaar (N=78 vragen)	36 (14)	38 (14)	37 (13)	38 (14)
> 75 jaar (N=16 vragen)	47 (24)	53 (22)	38 (20)	40 (25)
niet leeftijdgebonden (N=17 vragen)	51 (21)	54 (18)	55 (19)	52 (20)

Voor de categorie vragen over 'chronische aandoeningen' was alleen de gemiddelde score van de huisartsen met relatief veel ouderen in hun praktijk (gemiddelde score 47%), significant verschillend van de gemiddelde score van huisartsen met praktijken met andere bijzondere kenmerken (gemiddelde score 39% op deze categorie).

De gemiddelde kennisscores van huisartsen met praktijken met veel ouderen, verschilde voor de categorie vragen over 'ouderen (boven 75 jaar)' statistisch significant van de gemiddelde scores van huisartsen met praktijken met andere speciale kenmerken. De huisartsen die aangaven relatief veel ouderen in hun praktijk te hebben, bereikten op deze categorie een gemiddelde score van 53%, tegenover een totaal gemiddelde van 46% op deze categorie (zie tabel 5.4). De huisartsen met veel jonge gezinnen in de praktijk, behaalden gemiddeld op de vragen over oudere patiënten, een significant verschillende (lagere) score (38%) dan met name de huisartsen met een doorsnee-praktijk en de huisartsen met veel ouderen in hun praktijk.

De groep huisartsen met praktijken gekenmerkt door relatief veel jonge gezinnen, behaalde wel de hoogste gemiddelde scores (61%) op de toetsvragen in de categorie 'leeftijd patiënt jonger dan 15 jaar'. De verschillen met de gemiddelde scores van de overige subgroepen huisartsen waren echter, behalve voor de huisartsen met het praktijkkenmerk 'overig bijzonder' (waaronder bijvoorbeeld praktijken met veel allochtonen), niet significant.

In tabel 5.7 zijn de correlaties weergegeven van de 'algemene kennis' en een aantal numerieke achtergrondvariabelen van de huisartsen.

Tabel 5.7. *Samenhang numeriek verdeelde achtergrondvariabelen van huisartsen met hun kennisniveau, uitgedrukt in gemiddelde goed-min-foutscore (45%, std 11%) op de algemene kennistoets.*

Achtergrondvariabelen	Aantal	Resultaat correlatie berekeningen
<i>Leeftijd</i> gemiddeld 44 jaar (std 5.9)	340	Pearson corr. coëff.: - 0.27 Pr>R: 0.0001*
<i>Werkervaring</i> gemiddeld 13.8 jaar (std 6.3)	351	Pearson corr. coëff.: - 0.23 Pr>R: 0.0001*
<i>Werktijdfactor</i> gemiddeld 4.7 dagen per week (std 0.6)	353	Pearson corr. coëff.: - 0.04 Pr>R: 0.5105
<i>Duur opleiderservaring</i> gemiddeld 6.9 jaar (std 4.6)	280	Pearson corr. coëff.: - 0.05 Pr>R: 0.3970
<i>Cursorische nascholing</i> gemiddeld 14 dagdelen per jaar (std 10)	349	Pearson corr. coëff.: 0.01 Pr>R: 0.7861
<i>Bijhouden vakliteratuur</i> gemiddeld 2.7 uur (std 1.9)	349	Pearson corr. coëff.: 0.10 Pr>R: 0.0742
<i>Patiëntencontacten</i> gemiddeld 172 per week (std 58.6)	353	Pearson corr. coëff.: - 0.00 Pr>R: 0.9417
<i>Praktijkgrootte</i> gemiddeld: 2818 patiënten (std 953.2)	353	Pearson corr. coëff.: 0.03 Pr>R: 0.5681

* significant verband

Leeftijd en werkervaring correleren negatief met de kennistoetsscore. Oudere huisartsen, c.q. huisartsen met meer werkervaring, behalen gemiddeld een lagere score op de toets dan jongere huisartsen. Het leeftijdseffect blijft bestaan, ook als wordt gecontroleerd voor 'werkervaring' en het al dan niet gevolgd hebben van de huisartsopleiding.

Noch de werktijdfactor, noch de duur van de opleiderservaring, noch de mate van nascholingsactiviteiten (zoals uitgedrukt in de hoeveelheid cursorische nascholing die jaarlijks wordt bijgewoond of het bijhouden van de vakliteratuur), noch praktijkgrootte of aantal patiëntencontacten, correleren met de gemiddelde algemene kennistoetsscore van huisartsen.

De samenhang van de kennistoetsscores van de ervaren huisartsen met het oordeel dat zij gaven over de relevantie en toepasbaarheid van de kennistoets wordt weergegeven in tabel 5.8. De satisfactiescores zijn gebaseerd op de gemiddelde resultaten op een vijfpuntsschaal (1 = zeer eens; 5 = zeer oneens).

Tabel 5.8. *Samenhang tussen satisfactie huisartsen ten aanzien van de kennistoets en algemeen kennisniveau (N = 353 huisartsen).*

Satisfactievariabelen	Resultaat correlatie berekeningen
<i>Prettig toets te maken</i> gemiddelde score: 1.97 (std 0.84)	Pearson corr. coëff.: - 0.08 Pr>R: 0.1414
<i>Onderwerpen relevant</i> gemiddelde score: 1.84 (std 0.69)	Pearson corr. coëff.: - 0.01 Pr>R: 0.8859
<i>Toets is te lang</i> gemiddelde score: 3.12 (std 1.14)	Pearson corr. coëff.: 0.05 Pr>R: 0.3432
<i>Vragen relevant</i> gemiddelde score: 2.00 (std 0.74)	Pearson corr. coëff.: 0.03 Pr>R: 0.5262
<i>Juist-onjuistvorm is lastig</i> gemiddelde score: 3.64 (std 1.13)	Pearson corr. coëff.: 0.08 Pr>R: 0.1243
<i>Toets bruikbaar voor bepalen eigen kennislacunes</i> gemiddelde score: 2.11 (std 0.99)	Pearson corr. coëff.: - 0.04 Pr>R: 0.4917
<i>Toets bruikbaar bij nascholingskeuze</i> gemiddelde score: 2.44 (std 1.16)	Pearson corr. coëff.: - 0.08 Pr>R: 0.1644

Geen enkele satisfactievariabele vertoont een significante correlatie met de gemiddelde kennisscore. De huisartsen maken de toets gemiddeld niet beter of slechter naarmate hun oordeel over de toets of onderdelen danwel aspecten ervan, positiever of negatiever is. Ook het oordeel over de moeilijkheidsgraad van de juist-onjuist-vraagvorm hangt niet samen met de kennisscore die wordt behaald.

In een stapsgewijze multiële regressievergelijking is nagegaan welke combinatie van achtergrondkenmerken zoveel mogelijk verklaart van de variatie in kennistoetsscores. In de regressievergelijking werd de gemiddelde goed-min-foutscore van de 353 huisartsen gehanteerd als afhankelijke variabele. Het resultaat wordt weergegeven in tabel 5.9.

Tabel 5.9. *Resultaat stapsgewijze multiële regressievergelijking met gemiddelde goed-min-foutscore als afhankelijke variabele (N=353 huisartsen).*

Onafhankelijke variabelen	Proportie verklaarde variantie
Lidmaatschap NHG	0.0521
Leeftijd	0.0443
Aantal jaar werkervaring	0.0192
Lidmaatschap toetsingsgroep	0.0125
Totaal	0.1281

Het totale percentage verklaarde variantie bedraagt 13% bij een significantieniveau van 0.15. Bij dit niveau bleken alleen de combinatie van NHG-lidmaatschap, de leeftijd van de huisarts,

diens werkervaring en het al dan niet lid zijn van een toetsingsgroep enige voorspellende waarde te hebben ten aanzien van de kennistoetsscore.

5.4. CONCLUSIES KENNISNIVEAU EN ACHTERGRONDKENMERKEN HUISARTSEN

De groep huisartsen aan wie de kennistoets is voorgelegd, was geen representatieve steekproef uit de populatie van Nederlandse huisartsen. De groep van 353 huisartsen die in dit deel van het onderzoek is betrokken, week echter wat betreft achtergronden nauwelijks af van de Nederlandse populatie. De werving van deelnemers verliep relatief eenvoudig: vrijwel alle aangezochte groepen huisartsen waren bereid aan het onderzoek mee te doen. Uit het onderzoek komen geen aanwijzingen naar voren dat de groep deelnemers naar verhouding meer dan andere huisartsen in kennistoetsing geïnteresseerd of ervoor gemotiveerd was.

Uit de constructvalideringsstudie die is beschreven in hoofdstuk 2, was al duidelijk geworden dat de algemene huisartsgeneeskundige kennistoets in staat is om verschillen in gemiddelde kennis aan te tonen tussen huisartsen in diverse fasen van hun opleiding of ervaring. De gevonden verschillen tussen co-assistenten huisartsgeneeskunde, huisartsen-in-opleiding en ervaren huisartsen waren statistisch significant ($p < 0.05$).

De gemiddelde score van ervaren huisartsen op de kennistoets, ligt ongeveer op het niveau van die van aanstaande huisartsen na driekwart van hun tweejarige opleiding. Met toenemende ervaring dalen de gemiddelde scores op de kennistoets. Deze bevinding is conform de verwachting met betrekking tot het kennisbegrip dat in de toets aan de orde wordt gesteld. Huisartsen die de huisartsopleiding hebben gevolgd, hebben significant hogere scores dan huisartsen die dat niet hebben gedaan. Deze bevinding hangt echter direct samen met het leeftijdseffect.

Lidmaatschap van de wetenschappelijke beroepsvereniging NHG en opleiderservaring vertonen een significant positief verband met de kennistoetsscores. Deze bevinding is in de literatuur vaker gesignaleerd en wijst er mogelijk op dat deze groepen huisartsen actiever en beter geïnformeerd zijn dan de overige huisartsen. In tegenstelling tot wat op basis van de literatuur werd verwacht, is geen verband aangetroffen met werkverband of praktijksituering en kennisniveau.

De stapsgewijze multipele regressie-analyse liet zien dat het NHG-lidmaatschap, de leeftijd en werkervaring en het lidmaatschap van een toetsingsgroep een, zij het geringe, bijdrage leverden aan de verklaarde variantie van de toetsscore van de huisartsen. Het totale percentage verklaarde variantie bedroeg 13. Dat betekent dat slechts 13% van de variantie van de kennisscore kan worden verklaard uit de combinatie van deze vier onafhankelijke variabelen. De overige achtergrondvariabelen voegden daaraan niets meer toe. Het overgrote deel van de variantie in algemene kennisscore kan dus niet worden verklaard uit verschillen in achtergrondkenmerken van de huisartsen.

Bij conclusies over de kennisbeheersing op onderwerpen uit de algemene kennistoets passen kanttekeningen. De betrouwbaarheid van de gemiddelde toetsscores is beschreven in hoofdstuk 3. De conclusie was dat gemiddelde scores al bij relatief kleine groepen betrouwbaar zijn, en dat de betrouwbaarheid op onderdelen van de toets bij een voldoende grote groep deelnemers,

eveneens adequaat is. De onbetrouwbaarheidsmarges zijn groter naarmate een onderdeel uit minder vragen bestaat en/of scores voor een minder grote groep huisartsen worden berekend. Bij de interpretatie van de resultaten moet dit gegeven worden betrokken. Sommige onderwerpen waren blijkbaar aan de gemakkelijke kant omdat de co-assistenten er al relatief hoge scores op behaalden. Dit geldt bijvoorbeeld voor de onderwerpen 'bloed' (weliswaar slechts bestaande uit twee vragen) en 'oor' (5 vragen). Het onderwerp 'theorie' (huisartsgeneeskundig-theoretische aspecten) komt blijkbaar ook in de basisopleiding al uitgebreid aan bod. De gemiddelde score van ervaren huisartsen wijkt voor dit onderwerp niet af van de gemiddelde scores die op andere deelgebieden worden aangetroffen.

Interessanter dan de absolute hoogte van de gemiddelde scores per onderwerp, zijn de relatieve verschillen in het verloop van de kennis voor de verschillende deelnemersgroepen. Op onderdelen van de toets vertonen de gemiddelde scores van de co-assistenten, huisartsen-in-opleiding en ervaren huisartsen verschillende patronen. Het verloop van de scores voor sommige klachten- en aandoeninggebonden onderwerpen doet vermoeden dat de kennisontwikkeling daarvan nauw verbonden is met de mate waarin huisartsen er in de praktijk mee worden geconfronteerd. Dit geldt bijvoorbeeld aandoeningen van het bewegingsapparaat of huidaandoeningen. Voor dergelijke onderwerpen geldt bovendien dat de ervaren huisartsen, in tegenstelling tot het dalende patroon dat voor de totale toets wordt gezien, met toenemende ervaring gemiddeld hetzelfde kennisniveau behouden. Huidziekten en aandoeningen van het bewegingsapparaat staan hoog genoteerd op de 'top 10' van aandoeningen die huisartsen in hun praktijk tegenkomen (de Haan e.a., 1992; NIVEL, 1994).

Als wordt gekeken naar scoreverschillen tussen de deelnemersgroepen over verschillende fasen van het consult, valt voor de ervaren huisartsen op dat de gemiddelde scores op het gebied van diagnostiek en beleid (medicamenteus danwel niet medicamenteus) elkaar nauwelijks ontlopen. Voor alle groepen die nog in opleiding zijn (dat wil zeggen co-assistenten en aanstaande huisartsen in elke opleidingsfase), geldt echter dat de gemiddelde scores voor medicamenteus beleid achterblijven bij die voor diagnostiek en niet-medicamenteus beleid. Medicamenteus beleid betreft blijkbaar een gebied dat in de praktijk moet worden geleerd. Deze bevinding spoot met de bevindingen uit het onderzoek van Tan (1989), die laat zien dat de behoefte aan scholing van huisartsen-in-opleiding ten aanzien van diagnostiek geringer is dan de behoefte aan scholing ten aanzien van behandeling en beleid, en met het onderzoek van Zuidweg (1994) naar de kennis over diabetes in de huisartsopleiding.

Vergelijking van de gemiddelde scores op deelgebieden uit de toets, van huisartsen met praktijken die worden gekenmerkt door relatief veel ouderen, liet een significant verband zien tussen de scores op de deelgebieden 'ouderen (boven 75 jaar)' en 'chronische aandoeningen' voor huisartsen met praktijken die worden gekarakteriseerd door de aanwezigheid van relatief veel ouderen. Huisartsen met veel jonge gezinnen in de praktijk scoren gemiddeld wel hoger op de categorie 'jongeren beneden 15 jaar' maar significant is het verschil met de andere huisartsen niet. Huisartsen met praktijken met veel jonge gezinnen scoren wel significant lager op de categorie 'patiënten ouder dan 75 jaar'. Deze bevindingen sporen met de verwachtingen en wijzen erop dat variatie in kennis bij ervaren huisartsen waarschijnlijk nauw samenhangt met bepaalde kenmerken van de praktijk die ertoe leiden dat huisartsen relatief meer of minder met specifieke problematiek van patiënten in aanraking komen.

Hoofdstuk 6

ZELFKENNIS ERVAREN HUISARTSEN

SAMENVATTING

Om de relatie vast te stellen tussen het kennisniveau van huisartsen en hun eigen oordeel over hun expertise, werden de toetsscores van de huisartsen vergeleken met het oordeel dat ze over hun eigen expertise op een aantal terreinen gaven.

Ruim 40% van de ervaren huisartsen beschouwt zichzelf als deskundiger dan collega's op het gebied van psychosociale aandoeningen. Van de vrouwelijke huisartsen geeft zelfs meer dan 70% aan zich op dit terrein specifiek deskundig te achten. De onderwerpen 'neurologie', 'bloed' en 'mannelijke geslachtsorganen' worden daarentegen door minder dan 5% van de ervaren huisartsen aangewezen als speciaal expertisegebied. Vrijwel alle huisartsen beschouwen zichzelf als minstens even deskundig als collega's op de onderwerpen 'tractus digestivus', 'tractus respiratorius' of 'urinewegen'.

Gemiddeld gaven de huisartsen meer dan twee gebieden aan waarop zij zichzelf deskundiger achtten dan collega's. Er kon geen samenhang worden vastgesteld tussen het aantal opgegeven expertisegebieden en de gemiddelde toetsscore. Wel behaalde in het algemeen de groep huisartsen die aangaf op een bepaald onderwerp 'expert' te zijn, de hoogste kennistoetsscores op dat gebied. Deze verbanden waren maar voor een beperkt aantal onderwerpen statistisch significant. Een omgekeerde relatie werd wel gevonden.

De grote groep huisartsen die zich specifiek deskundig achtte op het gebied van psychosociale aandoeningen, behaalde op dit terrein geen significant hogere scores dan de overige huisartsen. Ook de vrouwelijke huisartsen behaalden geen hogere scores dan hun mannelijke collega's op dit specifieke gebied. Omgekeerd gaf bijna een kwart van de huisartsen aan zich geen expert te voelen op endocrien gebied. Deze groep behaalde geen significant lagere kennistoetsscores op dat gebied.

De conclusie is dat er geen eenduidige relatie bestaat tussen de kennistoetsscore en de inschatting van de eigen expertise. Het lijkt het erop dat de zelfkennis van huisartsen zowel onderwerp- als sexegebonden verschillen vertoont.

6.1. INLEIDING

De samenhang tussen het inzicht in de eigen expertise en het objectief bepaalde kennisniveau is van belang, omdat het tot nog toe in Nederland gebruikelijk is dat de huisartsen zelf inschatten wat ze aan nieuwe kennis nodig hebben en daarop hun nascholing afstemmen. De samenleving verwacht van artsen dat ze beschikken de vaardigheid de eigen (on)vermogens goed in te schatten (Gordon, 1992). Vanuit educatief perspectief wordt een toenemend belang

gehecht aan vormen van zelftoetsing, zeker in het kader van het 'levenslang blijven leren' (Boud, 1989). In situaties waarbij artsen geheel vrij zijn zelf hun nascholingsprogramma te bepalen, is deze vaardigheid van extra belang (Suter et al, 1984). In het algemeen lijkt er een overmaat aan zelfvertrouwen bij artsen te bestaan (Gordon, 1991). Dat zou betekenen dat niet altijd wordt gekozen voor nascholing op thema's, waar objectief bezien, de meeste aandacht naar uit zou moeten gaan. In de literatuur wordt er inderdaad op gewezen dat verschillen kunnen worden gesignaleerd tussen objectief bepaalde en subjectief ervaren leerbehoeften (Sibley et al, 1982; Grol & Zwaard, 1990).

In dit hoofdstuk wordt ingegaan op de samenhang tussen de kennis van huisartsen, zoals gemeten met de algemene huisartsgeneeskundige kennistoets, en het eigen oordeel van de huisartsen die aan deze toets deelnamen, over hun expertise op de medisch-inhoudelijke gebieden die in de toets naar voren kwamen.

Onderzoeksvraag II.3.:

In hoeverre hangt het oordeel dat huisartsen zelf aangeven over hun specifieke expertise samen met hun kennisniveau?

Opzet onderzoek:

Bepalen samenhang tussen kennistoetsscore ervaren huisartsen met antwoord op vraag naar inschatting eigen specifieke deskundigheid

6.2. METHODE BEPALING SAMENHANG KENNIS EN OORDEEL OVER EIGEN EXPERTISE

In de vragenlijst naar achtergrondkenmerken die de deelnemers aan de kennistoets van februari 1992 beantwoordden, was een vraag opgenomen naar de inschatting van de eigen expertise op een aantal medisch-inhoudelijke deelgebieden. Het betreft dezelfde gebieden die ook in de kennistoets aan bod kwamen. De antwoorden van de huisartsen op deze vraag konden zo in verband worden gebracht met hun kennistoetsscores.

In de vraagstelling met betrekking tot de eigen expertise werd gekozen voor een benadering waarbij de huisarts werd gevraagd de eigen expertise in te schatten in relatie tot de expertise van collega-huisartsen. Voor deze benadering werd gekozen omdat werd verwacht dat de relatieve benadering het meest aansluit bij de manier waarop huisartsen in de praktijk hun eigen deskundigheid inschatten. De vraag werd als volgt geformuleerd (zie ook Bijlage 6):

EXPERTISEGEBIEDEN

Elke huisarts heeft zo zijn of haar eigen speciale aandachtsgebieden, belangstellingen of stokpaardjes. De een kan bijvoorbeeld veel weten van diabetes, de ander gebruikt speciale apparatuur, zoals ECG's of oogheelkundige apparatuur. Wij vragen u hieronder om per onderwerp aan te geven op welke gebieden uw expertise naar uw eigen idee groter, ongeveer gelijk, dan wel kleiner is dan die van uw collega's.

Onderwerpen	relatieve expertise		
	veel	gemiddeld	weinig
Klachten en aandoeningen op het gebied van:			
a het bloed en bloedvormende organen			
b de tractus digestivus			
c het oog			
d het oor			
e de tractus circulatorius			
f het bewegingsapparaat			
g het zenuwstelsel			
h psychosociale klachten en aandoeningen			
i de tractus respiratorius			
j de huid			
k endocriene klieren, metabolisme, voeding			
i de urinewegen			
m zwangerschap, bevalling, anticonceptie			
n de vrouwelijke geslachtsorganen (incl. mammae)			
o de mannelijke geslachtsorganen			

Er werd van afgezien te vragen naar de categorie 'algemeen' en de categorie 'huisartsgeneeskundig-theoretische onderwerpen', omdat werd verwacht dat de vraag naar eigen expertise van de huisartsen in deze categorieën nauwelijks te beantwoorden zou zijn. De aansluiting aan de nascholingspraktijk van deze onderwerpen is bovendien gering.

Als kennisscores zijn steeds de gemiddelde goed-min-foutscores gehanteerd. Voor het verband tussen de totale toetsscore en het aantal genoemde expertisegebieden werd een correlatiecoëfficiënt (Pearson) berekend. Vervolgens zijn variantie-analyses gedaan per inhoudsgebied om het verband met de gemiddelde score over het betreffende gebied te kunnen beschrijven.

6.3. RESULTAAT SAMENHANG KENNIS EN OORDEEL OVER EIGEN EXPERTISE

Toetsscores en antwoorden op de expertisevraag waren beschikbaar van 347 huisartsen. Gemiddeld gaven zij 2.3 expertisegerieden aan. De antwoorden van de huisartsen op de vraag naar de inschatting van de eigen expertise, zijn samengevat in tabel 6.1.

Tabel 6.1. *Expertise zoals door huisartsen zelf aangegeven voor verschillende gebieden uit de kennistoets. (N=347 huisartsen)*

Onderwerpen	Expertise (%)			
	meer dan collega's	evenveel als collega's	minder dan collega's	totaal
Bloed	4	80	16	100
Tractus digestivus	10	89	1	100
Oog	12	69	19	100
Oor	6	91	3	100
Tractus circulatorius	23	74	3	100
Bewegingsapparaat	33	57	10	100
Neurologie	5	81	14	100
Psychische problemen	41	52	7	100
Tractus respiratorius	36	63	1	100
Huid	12	74	14	100
Endocrien/metab/voeding	6	70	24	100
Urinewegen	9	90	1	100
Zwangerschap etc.	28	54	19	100
Vrouwelijke geslachtsorganen	19	79	2	100
Mannelijke geslachtsorganen	3	91	6	100

Per onderwerp wordt duidelijk hoeveel huisartsen zich op het gebied meer, even of minder deskundig voelen dan collega's. Ruim 40% van de gevestigde huisartsen beschouwt zichzelf meer expert dan collega's op het terrein van psychische problemen. Ook op het gebied van de tractus respiratorius en op het terrein van het bewegingsapparaat achten relatief veel huisartsen zich deskundiger dan collega's. Bijna een kwart van de huisartsen beschouwt zichzelf minder deskundig dan collega's wat betreft aandoeningen en klachten op het gebied van de endocriene klieren, voeding en metabolisme. Wat betreft klachten en aandoeningen van het oog en zwangerschap/anticonceptie en bevalling beschouwt bijna twintig procent van de huisartsen zichzelf als minder deskundig dan collega's. Onderwerpen waarover vrijwel alle huisartsen aangeven gemiddeld of meer dan gemiddeld deskundig te zijn, zijn bij voorbeeld de tractus digestivus, de tractus respiratorius en de urinewegen.

Vervolgens is nagegaan of het type expertisegebieden dat door de huisartsen werd genoemd, verschilde voor huisartsen met verschillende achtergrondkenmerken. Hierbij bleek dat het beeld van vrouwelijke huisartsen van hun eigen expertise anders is dan het beeld dat mannelijke huisartsen van hun eigen expertise hebben. Zo gaf 72% van de 43 vrouwelijke huisartsen aan zich meer dan collega's deskundig te achten op het gebied van psychische klachten en aandoeningen, tegenover 37% van de mannelijke huisartsen. Wat betreft klachten

en aandoeningen van de vrouwelijke geslachtsorganen acht 70% van de vrouwelijke huisartsen zich expert tegen 12% van de mannelijke huisartsen. Ook op het gebied van zwangerschap, anticonceptie en bevalling achten relatief meer vrouwelijke huisartsen zich deskundig dan mannelijke huisartsen (57% versus 23%).

Daarentegen achten mannelijke huisartsen zich relatief vaker deskundig dan hun vrouwelijke collega's op het gebied van het bewegingsapparaat (35% tegenover 19%) en de tractus circulatorius (25% versus 12%). Voor de overige onderwerpen werden geen verschillen gevonden of was het aantal huisartsen dat zich expert op het betreffende gebied noemde, te klein om er verdere conclusies aan te verbinden (Pollemans et al, 1993).

Zoals eerder vermeld, werden gemiddeld 2.3 expertisegerieden aangegeven door de ervaren huisartsen. De gemiddelde goed-min-foutscore van de huisartsen bedroeg 45%. Er werd *geen* samenhang aangetroffen tussen het totale aantal genoemde expertisegerieden en de gemiddelde toetsscore (Pearson correlatie-coëfficiënt bedroeg 0.09). Met andere woorden, huisartsen behaalden, naarmate zij zichzelf op een groter aantal gebieden speciaal deskundig achtten, gemiddeld geen hogere kennistoetsscore.

Het verband tussen de zelf aangegeven expertise op bepaalde onderwerpen en de kennisscore op datzelfde onderwerp is weergegeven in tabel 6.2.

Tabel 6.2. Gemiddelde score op onderwerpen algemene kennistoets van huisartsen/deelnemers kennistoets februari 1992. Huisartsen ingedeeld naar zelf aangegeven expertiseniveau.

Onderwerpen	Gemiddelde goed-min-foutscore op onderwerpen algemene kennistoets (%)			
	totale groep huisartsen (N=347)	'expert' groepen	groepen met normale expertise	'niet-expert' groepen
Bloed	66	64	67	59
Tractus digestivus	48	52	47	44
Oog*	26	44	25	14
Oor	51	51	50	68
Tractus circulatorius	51	53	50	50
Bewegingsapparaat	37	40	36	36
Neurologie	23	33	24	20
Psychische problemen	35	38	34	29
Tractus respiratorius*	62	64	61	43
Huid*	59	69	59	50
Endocrien/metab/voeding	58	63	58	57
Urinewegen	38	36	38	40
Zwangerschap etc.	31	35	32	27
Vrouwelijke geslachtsorganen	16	26	14	19
Mannelijke geslachtsorganen	44	67	43	43

* verschillen significant

Voor vrijwel alle onderwerpen geldt dat de gemiddeld behaalde score door de experts de hoogste is, en dat de gemiddelde score van de niet-experts de laagste is. Alleen de thema's

'oor' en 'urinewegen' gedragen zich in dit opzicht anders. De verschillen zijn overigens alleen significant voor de thema's 'oog', 'tractus respiratorius' en 'huid'.

Gecontroleerd is tenslotte of de 'sexe-specifieke' bevindingen bij de inschatting van de eigen expertise, zich ook zouden voordoen in het scorepatroon van mannelijke en vrouwelijke huisartsen. De relatief grote groep vrouwelijke huisartsen die zich speciaal deskundig noemde op het gebied van psychische aandoeningen, behaalde gemiddeld geen hogere score op de toetsvragen op dit terrein. De vrouwelijke huisartsen behaalden daarentegen wel hogere scores dan hun mannelijke collega's op het gebied van vrouwelijke geslachtsorganen en op het gebied van zwangerschap, anticonceptie en bevalling. Voor sommige onderwerpen werd een met het expertise-oordeel strijdige score behaald. Zo scoorde de groep vrouwelijke huisartsen hoger dan mannelijke collega's op het gebied van de tractus circulatorius, terwijl relatief twee maal zoveel mannelijke huisartsen dan vrouwelijke collega's zich op dit gebied specifiek deskundig noemde.

6.4. CONCLUSIE ZELFKENNIS HUISARTSEN

De conclusie is dat de relatie tussen kennisscore en inschatting van eigen expertise niet eenduidig is. In algemene zin scoren huisartsen nauwelijks hoger op de kennistoets naarmate ze zich op meer gebieden deskundig achten. Als wordt gekeken naar de specifieke onderwerpen, scoren de huisartsen die zich expert noemen op bepaalde onderwerpen meestal op die bepaalde gebieden wél iets hoger dan hun collega's die zich geen expert voelen. De verschillen zijn echter vaak niet significant en het verband ligt ook wel eens andersom.

Afgemeten aan de samenhang tussen inschatting van eigen expertise en kennistoetsscores lijkt het erop dat de zelfkennis van huisartsen zowel onderwerp- als sexe-gebonden verschillen vertoont. Hoewel hier natuurlijk de kanttekening past dat expertise meer is dan een score op een algemene kennistoets, leveren de toetsscores daarmee toch een wat genuanceerder beeld op over de kennis van huisartsen op bepaalde onderwerpen, dan het eigen oordeel over de eigen deskundigheid.

Sommige onderwerpen springen eruit wat betreft het aantal huisartsen dat aangaf over specifieke expertise te beschikken: zo beschouwt meer dan 40% van de ervaren huisartsen zichzelf als deskundiger dan collega's op het gebied van psychosociale aandoeningen. Van de vrouwelijke huisartsen die in het onderzoek waren betrokken, geeft zelfs meer dan 70% aan zich op dit terrein specifiek deskundig te achten. De onderwerpen 'neurologie', 'bloed' en 'mannelijke geslachtsorganen' worden door minder dan 5% van de ervaren huisartsen aangewezen als speciaal expertisegebied. Bijna een kwart van de ervaren huisartsen beschouwt zichzelf niet als expert op het gebied van de 'endocriene klieren/metabolisme en voeding'. Er worden nauwelijks huisartsen aangetroffen die zichzelf als minder deskundig dan collega's beschouwen voor de onderwerpen 'tractus digestivus', tractus respiratorius' of 'urinewegen'.

Uit het onderzoek blijkt dat huisartsen niet over meer kennis beschikten naarmate zij zelf meer expertise-gebieden aangaven. Vergelijking van de scores van de groep die aangaf op een bepaald onderwerp 'expert' te zijn, met de scores van de groepen die aangaven over een gemiddelde of een minder dan gemiddelde expertise te beschikken, maakt duidelijk dat voor de meeste onderwerpen de groep die aangaf over relatief veel deskundigheid te beschikken,

ook de hoogste scores behaalde. De verbanden zijn echter maar voor een beperkt aantal onderwerpen significant en er wordt ook wel eens een omgekeerde relatie gevonden. De grote groep huisartsen die zichzelf expert acht op het gebied van psychosociale aandoeningen, scoort niet significant hoger dan de overige huisartsen. Ook de vrouwelijke huisartsen behalen geen hogere scores dan hun mannelijke collega's op dit gebied. De relatief grote groep huisartsen die aangaf zich weinig deskundig te beschouwen op endocrien gebied, behaalt geen significant lagere kennistoetsscores op dat gebied.

De volgende kanttekeningen moeten worden geplaatst bij deze conclusies. De kennistoets is niet ontworpen om het verband tussen subjectieve inschatting van eigen expertise over een aantal onderwerpen en de kennistoetsscores over die onderwerpen te kunnen bepalen. De toets is immers zodanig samengesteld dat ze een optimale dekking van het huisartsgeneeskundige kennisdomein vormt en ook nog binnen een redelijke tijd te beantwoorden is. Dat heeft tot gevolg dat sommige onderwerpen maar met een beperkt aantal vragen in de toets zijn vertegenwoordigd, omdat hun belang ten opzichte van andere onderwerpen relatief gering is. Dit betekent dat met name voor 'kleine' onderwerpen de groepsgemiddelden een relatief grotere onnauwkeurigheidsmarge hebben dan onderwerpen die met meer vragen in de toets zijn vertegenwoordigd. De onnauwkeurigheid wordt overigens ook beïnvloed door de groepsgrootte: ook kleine onderwerpen kunnen, als de groep maar groot genoeg is, toch redelijk nauwkeurige groepsscores opleveren (zie hoofdstuk 3).

Een tweede kanttekening moet worden gemaakt bij de manier waarop zelfkennis is gemeten. Gevraagd is naar een inschatting van de eigen expertise ten opzichte van collega's. Deze relativering is in de formulering aangebracht om sociaal wenselijke antwoorden zoveel mogelijk te voorkómen. Door de eigen expertise te vergelijken met anderen, was het immers niet nodig een absoluut oordeel over het eigen (on)vermogen uit te spreken. Verwacht werd ook dat deze benadering meer in overeenstemming was met de manier waarop huisartsen, ook bij het kiezen voor bepaalde nascholings thema's te werk gaan. Mogelijk kiezen huisartsen eerder voor bepaalde nascholings thema's als ze denken achter te blijven bij collega's. Wellicht had echter een meer directe vraag over de eigen expertise, dus niet in vergelijking met andere huisartsen, een ander resultaat opgeleverd.

DEEL III (hoofdstuk 7, 8 en 9)

ONDERWERPGEBONDEN KENNISTOETSING

De doelstelling van het derde deel van het onderzoek, en de vraagstellingen die daaromheen zijn geformuleerd, luiden als volgt.

Doelstelling III:

(Exemplarische) ontwikkeling van onderwerpgebonden kennistoetsen die praktisch toepasbaar zijn en waarmee valide en betrouwbaar kennisverandering in samenhang met gerichte nascholing kan worden gemeten.

Vraagstellingen:

- III.1. Wat is de validiteit van onderwerpgebonden kennistoetsen die bij specifieke nascholingsprogramma's worden ontwikkeld?
- III.2. Wat is de betrouwbaarheid van onderwerpgebonden kennistoetsen die bij specifieke nascholingsprogramma's worden ontwikkeld?
- III.3. Wat is de praktische toepasbaarheid in termen van acceptabiliteit en uitvoerbaarheid, van voor de nascholing bestemde, onderwerpgebonden kennistoetsen?

Deskundigheidsbevordering van huisartsen wordt meer en meer opgezet vanuit een bepaalde visie op nascholing. Daarbij wordt ervan uitgegaan dat deskundigheidsbevordering uiteindelijk zou moeten leiden tot verandering van het praktijkgedrag. De aanpak en de methoden waarvoor bij deskundigheidsbevordering wordt gekozen, moeten de voorwaarden daarvoor creëren en alle relevante facetten (bijvoorbeeld kennis, kunde, attitude, praktijkvoering) omvatten (NHG, 1990). Deskundigheidsbevordering moet zich inhoudelijk richten op de introductie van nieuwe en relevante inzichten, op de recente stand van zaken op een bepaald terrein en op nieuwe technieken of richtlijnen die wetenschappelijk goed zijn gefundeerd. Daarnaast moet deskundigheidsbevordering zijn gericht op het ophalen van leerstof die ooit is aangeleerd. Idealiter bevat deze een combinatie van interventies, die zijn gericht op de beïnvloeding van het functioneren van de individuele huisarts en diens praktijkvoering. Het aanbod dient verder gevarieerd te zijn, gebaseerd op een duidelijke probleemanalyse, en plaats te vinden in kleine groepen. Organisatie en structuur van de deskundigheidsbevordering moet tenslotte aan bepaalde randvoorwaarden voldoen. Op grond van deze uitgangspunten is een model geformuleerd dat het proces van deskundigheids- en kwaliteitsbevordering uiteenzet in een aantal stappen: oriëntatie, inzicht, acceptatie en verandering.

Deskundigheidsbevorderende activiteiten zouden met dit model rekening moeten houden en aandacht besteden aan de onderscheiden stappen (Grol & Zwaard, 1990). Zo zou bijvoorbeeld informatie over eigen lacunes in kennis of vaardigheden een rol kunnen spelen bij het ontstaan van een beter inzicht in de eigen competentie, en daarmee de bereidheid tot gedragsverandering kunnen stimuleren.

In dit deel van het onderzoek staat de vraag centraal welke rol kennistoetsen kunnen spelen in de hiervoor beschreven deskundigheidsbevordering, met andere woorden hoe kennistoetsen gebruikt kunnen worden als hulpmiddel om de kennis op een speciaal gebied te toetsen en vervolgens om verandering in kennis na het volgen van nascholing te kunnen vaststellen. Er is expliciet voor gekozen toetsen te ontwikkelen voor nascholingsprogramma's die zijn afgestemd op NHG-Standaarden voor het huisartsgeneeskundig handelen. Standaarden worden door het NHG ten behoeve van het kwaliteitsbeleid van de beroepsgroep ontwikkeld. Doordat ze het LHV-Basistakenpakket nader concretiseren, moeten ze de huisarts houvast in het werk bieden, en onaanvaardbare verschillen in aanpak tussen huisartsen terugdringen. De Standaarden worden beschouwd als basis voor toetsing, nascholing, en voor het maken van werkafspraken met specialisten. De Standaarden zijn opgebouwd volgens hetzelfde stramien. Dat stramien bestaat uit een nadere omschrijving van de inhoud, een inleiding en motivering, voorgesteld beleid, wetenschappelijke verantwoording, verslag van de totstandkoming van de Standaard, een overzicht van randvoorwaarden, literatuurverwijzingen en een overzichtskaart (NHG, 1990; Rutten & Thomas, 1993). De deskundigheidsbevorderingsprogramma's (pakketten) die bij de Standaarden worden samengesteld, vormen de basis voor gerichte nascholing over de belangrijkste elementen uit de Standaard. De pakketten zijn landelijk toepasbare onderwijsprogramma's, die beogen volgens de principes van volwasseneneducatie de inhoud van de Standaarden op een praktijkrelevante wijze aan huisartsen over te dragen (Grol & Zwaard, 1990). De programma's zijn afgestemd op het hiervoor aangeduide 4-stappen-model van deskundigheidsbevordering en kwaliteitsbewaking. Het is de bedoeling dat de nascholingsprogramma's die op de deskundigheidsbevorderingspakketten zijn gebaseerd, worden voorzien van korte kennistoetsen. Hiermee wordt enerzijds beoogd de deelnemers meer inzicht te geven in bestaande lacunes, en anderzijds de organisatoren en planners van de nascholing meer inzicht te geven in de effectiviteit van de door hen ontwikkelde programma's (NHG, 1993).

Om pragmatische redenen (er kon worden aangesloten bij lopende activiteiten in andere kaders) werd bij de exemplarische toetsontwikkeling voor de volgende onderwerpen gekozen: 'vaginaal bloedverlies' en 'cholesterol'. Het ontwikkelen van een specifieke toets en het toepassen van deze toets bij een nascholingsprogramma over het thema 'vaginaal bloedverlies' kon worden geëffectueerd in samenwerking met de afdeling Kwaliteitsbevordering van het NHG, in de periode dat het deskundigheidsbevorderingspakket werd samengesteld bij de NHG-Standaard Vaginaal Bloedverlies. De inhoud van deze Standaard was op dat moment nog niet gepubliceerd (Meijer e.a., 1992).

Wat betreft de ontwikkeling van een onderwerpgebonden toets en de toepassing van deze toets bij een nascholingsprogramma over het thema 'cholesterol' kon worden samengewerkt met het 'Implementatieproject cholesterol standaard': een WOK-onderzoek naar de implementatie van de NHG-Standaard Cholesterol in de huisartspraktijk, dat werd uitgevoerd aan de Rijksuniversiteit Limburg. De NHG-standaard Cholesterol is verschenen in 1991 (van Binsbergen e.a., 1991).

De inhoudsvaliditeit en de constructvaliditeit van de onderwerpgebonden kennistoetsen worden beschreven in hoofdstuk 7; de betrouwbaarheid in hoofdstuk 8 en de praktische toepasbaarheid in hoofdstuk 9.

Hoofdstuk 7

VALIDITEIT ONDERWERPGEBONDEN KENNISTOETSEN

SAMENVATTING

In dit hoofdstuk wordt de validiteit beschreven van onderwerpgebonden kennistoetsen die zijn bedoeld voor gebruik in de nascholing van huisartsen. Op geleide van NHG-Standaarden zijn, exemplarisch, twee onderwerpgebonden kennistoetsen ontwikkeld, één over 'vaginaal bloedverlies' en één over 'cholesterol'. De inhoudsvaliditeit en de constructvaliditeit van beide toetsen wordt beschreven.

Inhoudsvaliditeit is nagestreefd door de inhoud van de toetsen zo goed mogelijk af te stemmen op de kenniselementen die door deskundigen op het betreffende gebied beoordeeld werden als datgene wat een praktizerend huisarts beschikbaar moet hebben aan parate kennis over het betreffende onderwerp. De constructvaliditeit van de toetsen is onderzocht door deze in parallelvormen voor te leggen aan huisartsen direct voorafgaand aan en meteen na afloop van nascholing over het betreffende thema. Voor het onderwerp 'cholesterol' werd bovendien een retentiemeting verricht ongeveer acht maanden nadat de nascholing plaatsvond. Verandering in kennisniveau werd beschouwd als ondersteuning van de constructvaliditeit van de toetsen.

Aan de toetsing en nascholing over vaginaal bloedverlies namen 42 ervaren huisartsen deel. Deze toetsing werd gebruikt als pilotstudie om te kunnen nagaan of onderwerpgebonden kennistoetsing uitvoerbaar was en of met de kennistoets inderdaad verschil in kennisniveau kon worden gemeten. Aan het programma over cholesterol namen 28 ervaren huisartsen en 20 huisartsen-in-opleiding deel. Een vergelijkingsgroep maakte alleen de voortoets en de retentietoets. Deze groep bestond uit 34 huisartsen en 22 huisartsen-in-opleiding. De resultaten van de toetsafnames lieten, noch voor de groepen huisartsen onderling, noch voor de groepen huisartsen-in-opleiding onderling, statistisch significante verschillen zien tussen de gemiddelde scores op de toets voorafgaand aan het onderwijsprogramma (de voortoets). De verschillen in gemiddelde score tussen de totale groep huisartsen enerzijds en de totale groep huisartsen-in-opleiding anderzijds, waren wel significant. De huisartsen-in-opleiding behaalden daarbij lagere scores. De gemiddelde score die de verschillende deelnemersgroepen na afloop van de nascholing bereikten, was identiek. Acht maanden nadat de nascholing had plaatsgevonden was de gemiddelde score van de groepen die de nascholing hadden gevolgd nog steeds significant hoger dan hun aanvankelijke score, al was de score in vergelijking met de score die werd bereikt onmiddellijk na de nascholing wel gedaald. De gemiddelde score van de vergelijkingsgroepen was niet veranderd. Een stapsgewijze multiële regressie-analyse die werd uitgevoerd om te kunnen aangeven welke variabelen vooral bijdroegen aan het verschil in gemiddelde score tussen de voortoets en de retentietoets, liet alleen een significante samenhang zien tussen de retentietoetsscore en de score die was

behaald op de voortoets, en het al dan niet bijgewoond hebben van het nascholingsprogramma.

Geconcludeerd wordt dat het goed mogelijk is inhoudsvalide onderwerpgebonden kennistoetsen te construeren waarmee op groepsniveau verschillen kunnen worden gemeten die samenhangen met deelname aan een nascholingsprogramma. Deze laatste bevinding ondersteunt de constructvaliditeit van de toetsen.

7.1. INLEIDING

Nascholing voor (huis)artsen omvat een uitgebreide en gevarieerde reeks van onderwijsactiviteiten voor ervaren beroepsbeoefenaren. Deze moeten ertoe leiden dat artsen hun competentie handhaven (Anonymus, 1987). Veel literatuur heeft betrekking op de vraag of dit doel met nascholing inderdaad wordt bereikt (Berg, 1979; Goldfinger, 1982; Sibley et al, 1982; Wergin et al, 1988; Beaudry, 1989; Davis et al. 1990). Recente reviewstudies (zie hoofdstuk 1) wijzen erop dat het doel bereikbaar is, als is voldaan aan bepaalde onderwijskundige en methodologische eisen. Deze eisen gelden zowel de opzet en uitvoering van de nascholing, als de evaluatie daarvan. In dit hoofdstuk staat de validiteit centraal van onderwerpgebonden kennistoetsen, die exemplarisch zijn ontwikkeld op basis van de inhoud van NHG-Standaarden. De vraag is of deze toetsen gevoelig genoeg zijn om kennisverschillen ten gevolge van nascholing over deze Standaarden, te meten. Beschreven wordt hoe de validiteit is onderzocht van twee exemplarisch ontwikkelde kennistoetsen over de nascholingsthema's 'vaginaal bloedverlies' en 'cholesterol'. Voor beide toetsen wordt ingegaan op de inhoudsvaliditeit en de constructvaliditeit. De vraag naar de geschiktheid van de onderwerpgebonden toetsen om effecten van nascholing te meten, wordt opgevat als een constructvaliditeitsvraag. Voor de constructvalidering wordt gebruik gemaakt van gegevens die zijn verzameld bij toetsafnames voorafgaand aan en na afloop van nascholing van huisartsen. De resultaten van het onderzoek worden beschouwd als ondersteunend voor de constructvaliditeit van de toetsen, als het mogelijk is met de toetsen effecten te meten in samenhang met nascholing. Daarmee is de constructvalidering tevens een vorm van evaluatie-onderzoek. De ontwikkeling van de toets over vaginaal bloedverlies is vooral beschouwd als een pilotstudie naar de ontwikkelingsmogelijkheid en uitvoerbaarheid van onderwerpgebonden toetsing.

Onderzoeksvraag III.1.:

Wat is de validiteit van onderwerpgebonden kennistoetsen die bij specifieke nascholingsprogramma's worden ontwikkeld?

Opzet validiteitsstudie:

- inhoudsvaliditeit: verantwoording instrumentontwikkeling
- constructvaliditeit: scoreverloop voor-, na- en retentietoets

7.2. INHOUDSVALIDERING

Resultaten op onderwerpgebonden toetsen moeten objectieve informatie bieden over kennisveranderingen onder invloed van gerichte nascholing (NHG, 1993; Rutten & Thomas, 1993). De toetsen moeten dus een goede representatie vormen van de inhoud van de

nascholing. Evenals bij de algemene kennistoets is de inhoudsvalideringsstudie van de onderwerpgebonden kennistoetsen er vooral op gericht de inhoud van de toetsen wat betreft hun representativiteit voor het te toetsen gebied, goed te onderbouwen. In beide kennistoetsen die als voorbeeld ten behoeve van dit onderzoeksdeel werden ontwikkeld, is om praktische redenen bij de keuze van de vraagvorm aangesloten bij de vraagvorm in de algemene huisartsgeneeskundige kennistoets. De toetsen bestonden dan ook uit praktijkrelevante, huisartsgeneeskundige casuïstiek met bijbehorende (gesloten) vragen van het juist-onjuist-vraagtekentype. Omdat dezelfde vraagvorm als in de algemene kennistoets werd gehanteerd, kon er gebruik worden gemaakt van dezelfde gegevensverwerkende apparatuur, en konden de resultaten met elkaar worden vergeleken.

7.2.1. Kennistoets vaginaal bloedverlies

De inhoud van de kennistoets moest de meest wezenlijke kenniselementen uit de NHG-Standaard 'Vaginaal Bloedverlies' (Meijer et al, 1992) weerspiegelen. De toetsvragen werden daartoe afgeleid van de belangrijkste leerdoelen die waren geformuleerd door de makers van het deskundigheidsbevorderingspakket behorend bij deze NHG-Standaard. Concept-toetsvragen werden opgesteld door een huisarts-staf lid van het NHG, die tevens de samenstelling van het deskundigheidsbevorderingspakket coördineerde. De inhoud van de vragen en de aansluiting ervan op de NHG-Standaard werden gecontroleerd door twee huisartsen/toetsdeskundigen van het landelijk Samenwerkingsverband. Vier leden (huisartsen) van de Werkgroep Kennis van het Samenwerkingsverband beoordeelden de vormtechnische kwaliteit van de toetsvragen en stelden zo nodig alternatieve formuleringen voor. De definitieve vormgeving van de toets geschiedde onder verantwoordelijkheid van het landelijk Samenwerkingsverband. De kennistoets werd verondersteld inhoudsvalide te zijn als de leerdoelen en de inhoud van het onderwijsprogramma volledig werden gedekt, en als de toetsvragen voldeden aan de vormtechnische eisen die worden gehanteerd door het Samenwerkingsverband.

De kennistoets 'vaginaal bloedverlies' bestond uiteindelijk uit 26 vragen, alle in de vorm van naar huisartsgeneeskundige casuïstiek verwijzende stellingen die met juist, onjuist of een vraagteken dienden te worden beantwoord. Volgens de inhoudsdeskundige was met de toets het kennisdomein zoals gedefinieerd in de NHG-Standaard en de daarop gebaseerde doelstellingen, uitputtend gedekt. Vormtechnisch voldeden de toetsvragen aan de eisen van het landelijk Samenwerkingsverband (zie Bijlage 10 voor een voorbeeldpagina uit de toets over vaginaal bloedverlies).

7.2.2. Kennistoets cholesterol

De kennistoets cholesterol moest het kennisdomein dat volgens de NHG-Standaard Cholesterol (van Binsbergen et al, 1991) relevant is voor de dagelijkse praktijkvoering van de huisarts, dekken. Om dit te kunnen bereiken stelde één van de samenstellers van de NHG-Standaard Cholesterol eerst een lijst op van relevante kenniselementen uit de NHG-Standaard en de wetenschappelijke verantwoording ervan (epidemiologische en pathofysiologische achtergrondkennis, en kennis die van direct praktisch belang is bij diagnostiek en therapie). Deze lijst werd vervolgens in onderling overleg met de onderzoekers in het kennisproject, waarvan dit proefschrift een onderdeel vormt, op basis van enkele praktische criteria ingekort. De criteria behelsden:

- het belang voor de praktiserende huisarts van parate kennis over het betrokken kenniselement;
- de ondubbelzinnigheid van het kenniselement, dat wil zeggen de mogelijkheid om over het betreffende kenniselement een kennistoetsvraag te maken.

De lijst werd in twee rondes voorgelegd aan de afzonderlijke leden van de werkgroep die verantwoordelijk was voor de samenstelling van de NHG-Standaard. In de eerste ronde werd de zes leden van deze werkgroep verzocht de elementen te beoordelen op hun waarde als parate kennis voor huisartsen. Tevens werd hen verzocht aan te geven welke relevante kenniselementen uit de NHG-Standaard ontbraken. De bijgestelde lijst werd in de tweede ronde nogmaals door de leden van de werkgroep becommentarieerd. In de uiteindelijke lijst werden alle elementen opgenomen die door de helft of meer van de werkgroepleden werd aangegeven als belangrijk kenniselement, waarover praktiserende huisartsen parate kennis zouden moeten hebben. Deze procedure leidde tot inkorting van de oorspronkelijke groslijst van 57 kenniselementen tot een lijst van 27 'key features'. De resulterende lijst vormde de basis voor de ontwikkeling van een toegespitste kennistoets over het onderwerp Cholesterol (zie Bijlage 11). De inhoudelijke kwaliteit van de toetsvragen werd mede beoordeeld door één van de samenstellers van de NHG-Standaard Cholesterol. De uiteindelijke toets werd becommentarieerd op vormtechnische aspecten door drie leden van de Werkgroep Kennis van het Samenwerkingsverband. Ook voor de cholesteroltoets werd een paralleltoets gemaakt ten behoeve van het evaluatieonderzoek; een inhoudelijk identieke toets waarin alleen de vraagvolgorde en de namen van de patiënten in de toetscasuïstiek werden veranderd.

De kennistoets over cholesterol bestond uit 30 juist-onjuist-vraagtekenvragen, geformuleerd rondom praktijkrelevante casuïstiek. Hiervan hadden 6 vragen betrekking op epidemiologische aspecten, 12 op diagnostische en nogmaals 12 op therapeutische aspecten. Het lukte om over alle 'key features' toetsvragen te formuleren. De vragen voldeden aan de vormtechnische eisen van het Samenwerkingsverband (zie Bijlage 12 voor een voorbeeldpagina uit de toets over cholesterol).

7.3. CONSTRUCTVALIDERING

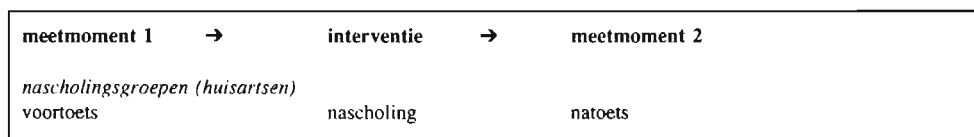
De constructvaliditeit werd onderzocht door het vaststellen van het verschil in gemiddelde score op inhoudelijk parallelle toetsen, afgenomen voorafgaand aan en (enige tijd) na deelname aan een nascholingsprogramma over het betreffende onderwerp. De literatuur geeft aan dat dit verschil de beste indicator is om groei in kennis op groepsniveau vast te stellen (Cronbach & Furby, 1970; Linn, 1988). Voor het onderwerp 'cholesterol' kon bovendien het verschil in scores van huisartsen-in-opleiding en ervaren huisartsen in de constructvalidering worden betrokken. De gevolgde procedures en groepen deelnemers die in dit deel van het validiteitsonderzoek werden betrokken, verschilden voor de twee onderwerpen. Daarom worden de constructvaliditeitsstudies voor beide onderwerpen in aparte paragrafen besproken. In de inleiding is gesteld dat de constructvalidering wordt opgevat als een vorm van onderwijsevaluatie-onderzoek. De basishypothese van elk evaluatie-onderzoek is een beleidsredenering, namelijk dat de uitvoering van het programma (ofwel de implementatie van het evaluatie-object) de gewenste effecten tot gevolg zal hebben (Hoeben, 1987). Dit is echter een causale redenering die nooit in een zuiver experimentele opzet kan worden onderzocht, omdat random-toewijzing van proefpersonen aan een experimentele of controlegroep in het onderwijs feitelijk onmogelijk is (Hofstee, 1987). Toch bestaat er wel overeenstemming over

de wijze waarop onderwijsevaluatie zou moeten plaatsvinden (Hoeben, 1987). Dat betekent onder meer dat controle op eventuele neveninvloeden op de effecten van het onderwijsprogramma, in de plaats komt van random toewijzing van proefpersonen. In het onderzoek wordt hierin op een aantal wijzen voorzien. In navolging van wat de literatuur suggereert, is er veel aandacht besteed aan de aansluiting en de wederzijdse dekking tussen onderwijs en toets (Leinhardt & Seewald, 1981). Daarbij wordt gecontroleerd of alle beoogde elementen in het onderwijsprogramma inderdaad adequaat aan bod zijn gekomen. In de constructvalideringsstudie over het thema 'cholesterol' is daarnaast een regressie-analyse uitgevoerd naar de invloed van verschillende relevante achtergrondvariabelen op de kennisontwikkeling.

7.3.1. Kennistoets vaginaal bloedverlies

- methode constructvalidering

Voor de constructvalideringsstudie van de onderwerpgebonden toets over 'vaginaal bloedverlies' was het *design* als volgt.



Dit design is een gemodificeerde versie van een eenvoudig pretest-posttest design, waarin in eerste instantie alleen wordt gekeken of er op een kennistoets die wordt afgenomen na afloop van een nascholingsprogramma hogere scores worden behaald dan op een toets voorafgaand aan de nascholing (Astin & Panos, 1971). Binnen het gekozen design werd informatie verzameld over de kenmerken van de deelnemers (scores op de voortoets en achtergrondkenmerken), over de interventie (de nascholing) en over de uitkomsten van de nascholing (de scores van de deelnemers op de natoets en satisfactiegegevens). Nadeel van het design is dat geen definitieve uitspraken mogen worden gedaan over de causaliteit van eventuele verbanden omdat er geen controlegroep in is opgenomen (Campbell, 1966).

Bij de constructvalidering werden de volgende *instrumenten* toegepast. Om de kennis van de deelnemers te toetsen, voorafgaand aan de nascholing werd de op de NHG-Standaard gebaseerde kennistoets over vaginaal bloedverlies gebruikt (zie Bijlage 10 voor een voorbeeldpagina uit de toets). De parallel-versie van deze toets werd direct na afloop van het nascholingsprogramma afgenomen. Evenals bij de algemene kennistoets werden de deelnemers vragen voorgelegd over persoonlijke en professionele achtergrondkenmerken. Hiervoor werd een aanpassing gemaakt van de lijst die in het onderzoek naar de algemene kennistoets was gebruikt (zie Bijlage 6). De vragenlijst was evenals de antwoordformulieren op de kennistoets ten behoeve van de anonieme verwerking van de gegevens, voorgecodeerd. Daarnaast werd een satisfactieformulier door de deelnemers ingevuld. Daarin werd onder meer gevraagd of de toetsen naar hun oordeel een goede afspiegeling vormden van de inhoud van de nascholing (zie Bijlage 13).

De *deelnemers* aan de nascholing en toetsing waren allen ervaren huisartsen. De werving kwam tot stand via afspraken van de afdeling kwaliteitsbevordering van het NHG met twee nascholingsgroepen. Inschrijving geschiedde, conform gebruik binnen deze groepen, vrijwillig.

De huisartsen werden tevoren geïnformeerd over de aan de nascholing verbonden toetsing (zie Bijlage 14). Alle deelnemers werd individuele, schriftelijke feedback over de eigen toetsresultaten, respectievelijk over de gemiddelde toetsresultaten van hun groepsgenoten en van alle deelnemende huisartsen toegezegd, alsmede een geringe vergoeding voor het meedoen aan de toetsingsactiviteiten. Het aantal deelnemers per toets, uitgesplitst naar groep, is in tabel 7.1 weergegeven.

Tabel 7.1. Deelnemers kennistoetsing vaginaal bloedverlies.

Deelnemers	Aantal	
	voortoets	natoets
<i>Nascholingsgroepen</i>		
groep 1	20	20
groep 2	22	22
Totaal	42	42

In totaal hebben 42 ervaren huisartsen aan het 'complete programma' van voortoets, nascholingsavond en natoets deelgenomen. Omdat de inschrijving op het programma vrijwillig was, is over 'non-respons' geen duidelijke uitspraak te doen. Tevoren was voor beide groepen alleen bekend wat de gemiddelde opkomst bij nascholingsavonden was. Voor groep 1 was de opkomst bij het nascholingsprogramma 'vaginaal bloedverlies' ongeveer gelijk aan de gebruikelijke opkomst, voor groep 2 gold dat de opkomst ruim boven gebruikelijk was.

Tabel 7.2. Achtergrondkenmerken deelnemers kennistoetsing vaginaal bloedverlies.

Achtergrondkenmerken	Deelnemers kennistoetsing vaginaal bloedverlies (N=42)
<i>Geslacht</i>	
man	62%
vrouw	38%
<i>Gemiddelde leeftijd</i>	40 jaar (std 6)
<i>Praktijksituering</i>	
grote stad	17%
stad	38%
versted. platteland	38%
platteland	2%
onbekend	5%
<i>Werkverband</i>	
solist	38%
duo/associatie	29%
groepspraktijk/gezondh.centrum	19%
overig/onbekend	14%

In tabel 7.2 zijn de belangrijkste achtergrondkenmerken weergegeven van de deelnemers. In vergelijking met het aantal vrouwelijke huisartsen in de Nederlandse huisartspopulatie (12%, NIVEL, 1992), was het aantal vrouwelijke huisartsen in de groep die aan de nascholing plus toetsing over vaginaal bloedverlies deelnam, groot. Dit werd veroorzaakt doordat een van beide nascholingsgroepen voor bijna tweederde deel uit vrouwen bestond.

De *interventie* bestond uit een voor landelijk gebruik bestemd nascholingsprogramma voor ervaren huisartsen, dat op de NHG-Standaard 'Vaginaal Bloedverlies' was gebaseerd. Met het programma werd beoogd de belangrijkste kenniselementen uit de NHG-Standaard over te dragen. In het nascholingsprogramma kwamen dezelfde kennisaspecten terug als waarop de kennistoets was toegespitst. Door het NHG was docenten- en cursistenmateriaal samengesteld en was een structuur voor de nascholingsbijeenkomst gegeven. Het docentenmateriaal bestond uit overhead-sheets met daarop in het kort de meest wezenlijke inhoudelijke aspecten uit de NHG-Standaard, en verder uit uitgewerkte casuïstiek. Voor de deelnemers waren er niet-uitgewerkte casusbeschrijvingen beschikbaar. De deelnemers beantwoordden eerst individueel enkele open vragen over de casuïstiek. Vervolgens bespraken ze hun antwoorden in kleine groepjes. Daarna werd plenair de voor de beantwoording van de casus relevante informatie uit de NHG-Standaard besproken. Het programma voorzag erin dat door de beantwoording van alle vragen bij de verschillende casus, alle relevante kennisaspecten aan bod kwamen. Op basis van de informatie op sheets kon worden gediscussieerd over de motivatie voor bepaalde keuzes in de NHG-Standaard. Ook was er een medisch specialist (gynaecoloog) bij de nascholing aanwezig ter verduidelijking van eventuele problemen. Over de gewenste aard en inhoud van de bijdrage van de expert waren afspraken gemaakt. Het programma 'vaginaal bloedverlies' nam één dagdeel (3 uur) in beslag.

De nascholing werd bijgewoond door de samensteller van het programma en door de kennistoets-onderzoeker. Door hen werd geïnventariseerd of de nascholing verliep volgens de vastgestelde planning, en of alle onderwerpen die in het programma aan bod moesten komen, ook daadwerkelijk en inhoudelijk correct aan bod kwamen.

Alle voorttoetsen werden thuis ingevuld. De deelnemers aan de nascholing kregen daartoe de benodigde toetsboekjes en (voorgecodeerde) antwoordformulieren enkele dagen voor de nascholingsbijeenkomst toegestuurd. Hen werd gevraagd de *toets* zonder overleg noch met gebruikmaking van naslagwerken in te vullen. De natoetsen werden groepsgewijs, onmiddellijk na afloop van de nascholing, zonder onderling overleg ingevuld.

Na afname werden de kennistoetsen eerst gecontroleerd op fouten, onduidelijke vragen of ontstane misverstanden. Op basis hiervan is één vraag uit de toets verwijderd. De definitieve toets waarover feedback werd gegeven en die in de nadere analyses is betrokken, bestond dus uit 25 vragen. De resultaten werden berekend als percentage goed, fout en met een vraagteken beantwoorde vragen (zie Bijlage 15).

- *resultaten constructvalidering*

In tabel 7.3 zijn de gemiddelde goed-min-foutcores weergegeven van alle deelnemers op de voor- en de natoets over vaginaal bloedverlies.

Tabel 7.3. Gemiddelde goed-min-foutscores ervaren huisartsen op voor- en natoets vaginaal bloedverlies (N=42 huisartsen).

Deelnemers	Gemiddelde goed-min-foutscore (%)	
	Voortoets	Natoets
<i>Nascholingsgroepen</i>		
nascholingsgroep 1 (N=20)	34 (std 25)	48 (std 16)
nascholingsgroep 2 (N=22)	28 (std 23)	75 (std 11)
Totaal (N=42)	31 (std 24)	62 (std 19)

Gemiddeld behaalden de huisartsen op de toets die voorafgaand aan de nascholing werd ingevuld, een goed-min-foutscore van 31%. Op de toets die onmiddellijk na afloop van de nascholing werd afgenomen was de gemiddelde score 62%.

De verschillen tussen de deelnemersgroepen op de voortoets zijn niet significant. De verschillen tussen de scores op de voor- en natoets zijn voor de afzonderlijke groepen en voor de totale groep wel significant ($p < 0.05$). Het grote verschil tussen beide nascholingsgroepen in de gemiddelde score op de natoets is opvallend. Het gebied waarin een goed-min-foutscore kan liggen varieert van +100 (alles goed) tot -100 (alles fout). De relatief grote standaarddeviaties worden veroorzaakt doordat de goed-min-foutscores van de deelnemers sterk uiteenliepen. Op de natoets zijn de standaarddeviaties van de gemiddelde scores minder groot dan op de voortoets. De scores van de deelnemers zijn dus dichter bij elkaar gekomen.

Tijdens de twee nascholingsbijeenkomsten werden in de uitvoering van het programma grote verschillen waargenomen. In nascholingsgroep 1 werd de tijdsplanning niet bewaakt, zodat een deel van het programma niet aan bod kwam. Noch de huisarts-docent noch de gynaecoloog die als expert achtergrondinformatie moest bieden, hield zich aan de afspraken over wat er in het programma inhoudelijk aan informatie zou worden geboden en op welke wijze dat zou gebeuren. Dat laatste leidde er onder meer toe dat, tijdens het programma, een uitgebreide inhoudelijke discussie ontstond over onderdelen van de NHG-Standaard. Over enkele aspecten werd verkeerde informatie geboden, dat wil zeggen tegenstrijdig aan de inhoud van de NHG-Standaard. In nascholingsgroep 2 vond daarentegen het programma plaats conform de afspraken. De docent in nascholingsgroep 2 was betrokken geweest bij de samenstelling van de betreffende NHG-Standaard. Deze docent was dan ook goed op de hoogte van de achtergronden van de inhoud van het onderwijsprogramma. Ook didactisch waren er grote verschillen tussen beide nascholingsbijeenkomsten. In nascholingsgroep 1 werden didactische onvolkomenheden signaleerd. De nascholing in deze groep bestond vooral uit het uitwisselen van persoonlijke ervaringen en opvattingen van de deelnemers. Door beide waarnemers aan de nascholingsbijeenkomsten werd dan ook geconcludeerd dat in nascholingsgroep 1 de kennisoverdracht onvolledig en op een aantal aspecten onjuist was. Omdat in beginsel het onderwijsprogramma en de toets naadloos op elkaar aansloten, kon worden geïnventariseerd welk deel van de stof, met betrekking tot welke toetsvragen niet, dan wel verkeerd tijdens de nascholing werd overgedragen. Het ging hierbij om in totaal 9 van de 26 vragen uit de kennistoets.

In tabel 7.4 worden de scores van de deelnemersgroepen onderscheiden naar het deel van de toets dat niet of gebrekkig aan bod is gekomen in nascholingsgroep 1 (de deelttoets 'lacunair onderwijs') en het onderwijsdeel dat wel in beide nascholingsgroepen aan bod is geweest (de deelttoets 'volledig onderwijs'). Het lacunaire deel betreft de gemiddelde scores over 9 toetsvragen, het volledige deel de gemiddelde scores over 16 toetsvragen. (Voor alle duidelijkheid: het 'lacunaire' deel is in nascholingsgroep 2 wel goed aan bod gekomen.)

Tabel 7.4. Toetsscores 'vaginaal bloedverlies' op voor- en natoets ten aanzien van lacunair versus volledig onderwezen kenniselementen (gemiddelde goed-min-foutscore per deelnemersgroep).

Deelnemers	Gemiddelde goed-min-foutscore (%)			
	in groep 1 'lacunair' onderwijs (9 vragen)		in beide groepen 'volledig' onderwijs (16 vragen)	
	voortoets	natoets	voortoets	natoets
Nascholingsgroepen				
groep 1 (N=20)	8 (std 38)	1 (std 30)	48 (std 26)	74 (std 17)
groep 2 (N=22)	2 (std 41)	67 (std 19)	43 (std 26)	80 (std 13)

Ook op de 'deelttoetsen' is op de voortoets het verschil tussen beide nascholingsgroepen niet significant. Het verschil in gemiddelde score op de natoets tussen beide nascholingsgroepen wat betreft de deelttoets 'lacunair onderwijs' is aanzienlijk. De goed-min-foutscore in nascholingsgroep 1 daalt van 8 op de voortoets naar 1% op de natoets. Voor nascholingsgroep 2 is het beeld geheel anders: hier stijgt de goed-min-foutscore van 2 naar 67%. Een dergelijk verschil wordt niet teruggevonden in de scores op de rest van de toets. Op het toetsdeel dat betrekking had op het onderwijs dat in beide groepen goed uit de verf kwam, is het scoreverloop in beide groepen ongeveer hetzelfde.

Op de natoets was het scoreverschil tussen beide nascholingsgroepen, op het deel van de toets dat het 'lacunaire' onderwijs in nascholingsgroep 1 weerspiegelde, wel significant. Op het resterende deel van de toets was het verschil tussen nascholingsgroep 1 en 2 niet significant. De toetsvragen die het deel van het nascholingsprogramma weerspiegelen dat in nascholingsgroep 1 niet goed aan bod is gekomen, lijken gemiddeld moeilijker te zijn dan de overige toetsvragen. Voor de interpretatie van het verschil in scores op voor- en natoets tussen de beide nascholingsgroepen, maakt dit echter geen verschil. Beide toetsdelen waren voor beide groepen respectievelijk moeilijker of gemakkelijker.

Bij het oordeel van de deelnemers aan de nascholing over de mate waarin onderwijs- en toetsinhoud met elkaar overeenkwamen, is het interessant om te kijken naar eventuele verschillen in opvattingen tussen de twee nascholingsgroepen (zie tabel 7.5).

Tabel 7.5. *Oordeel nascholingsdeelnemers over aansluiten kennistoets vaginaal bloedverlies op nascholingsprogramma.*

Opvattingen	Waardering (in percentages)		
	(geheel) eens	neutraal/geen mening	(geheel) oneens
<i>Inhoud voortoets is goede afspiegeling nascholingsprogramma</i>			
nascholingsgroep 1 (N=16)	56	31	13
nascholingsgroep 2 (N=13)	92	8	—
<i>Inhoud natoets is goede afspiegeling nascholingsprogramma</i>			
nascholingsgroep 1 (N=16)	44	38	19
nascholingsgroep 2 (N=13)	100	—	—

Niet alle deelnemers vulden de satisfactieformulieren (volledig) in. In nascholingsgroep 1 blijken de deelnemers die het formulier wel invulden, de kennistoetsen inderdaad als een minder goede afspiegeling van het nascholingsprogramma te beschouwen dan in nascholingsgroep 2, waar de toets en de inhoud van het programma, ook naar het oordeel van de waarnemers, elkaar veel beter dekten. De gegevens wijzen er niet op dat voor alle deelnemers duidelijk was dat de voor- en de natoets inhoudelijk identieke paralleltoetsen waren.

7.3.2. Kennistoets cholesterol

- methode constructvalidering

Het *design* van de studie waarin de constructvaliditeit van de kennistoets over cholesterol werd onderzocht, was als volgt.

meetmoment 1	→	interventie	→	meetmoment 2	→	interventie	→	meetmoment 3
<i>nascholingsgroepen (huisartsen en huisartsen-in-opleiding)</i>								
voortoets		nascholing		natoets		interventie in praktijk		retentietoets
voortoets		nascholing		natoets				retentietoets
<i>vergelijkingsgroepen (huisartsen en huisartsen-in-opleiding)</i>								
voortoets								retentietoets

De belangrijkste verschillen met het design bij de constructvalidering van de toets over vaginaal bloedverlies zijn de opname van vergelijkingsgroepen (waarbij een deel van de huisartsen op basis van een matchingsprocedure aan nascholings- danwel vergelijkingsgroep werd toebedeeld), de deelname van (een deel van) de huisartsen aan een tweede interventie die plaatsvond in de praktijk, en tenslotte de retentiemeting. De deelname aan de tweede interventie was gevolg van het feit dat voor het kennisonderzoek over het thema cholesterol werd aangesloten bij het implementatie-onderzoek naar de NHG-Standaard Cholesterol waarin een groep huisartsen na de nascholing over het onderwerp intensief in de praktijk werd begeleid, en een andere groep huisartsen niet (van der Weijden, 1992; van der Weijden, in voorbereiding).

Aan de nascholing en toetsing deden zowel groepen huisartsen, als groepen huisartsen-in-opleiding mee. De NHG-Standaard Cholesterol was op het moment van de toetsing en nascholing al enige jaren bekend. Daarom mocht worden verondersteld dat de ervaren huisartsen een gemiddeld hoger kennisniveau wat betreft dit thema zouden hebben dan de huisartsen-in-opleiding. Als de ervaren huisartsen gemiddeld een significant hogere score op de kennistoets zouden behalen dan de huisartsen-in-opleiding, werd dit beschouwd als ondersteuning voor de constructvaliditeit van de toets.

De *instrumenten* in dit deel van het onderzoek waren de onderwerpgebonden kennistoets over cholesterol, waarvoor een parallelversie werd gemaakt voor de natoets en vervolgens nogmaals voor de retentietoets (zie Bijlage 12 voor een voorbeeldpagina). Evenals bij de paralleltoets over vaginaal bloedverlies werd daartoe de vraagvolgorde gewijzigd en werden de namen en de leeftijden van de patiënten in de casus veranderd. De deelnemers aan de toetsen werd gevraagd een vergelijkbare vragenlijst naar achtergrondkenmerken in te vullen als die bij vaginaal bloedverlies. De deelnemers vulden een satisfactievragenlijst in bij de nameting, waarin werd gevraagd op een vijfpuntsschaal een oordeel te geven over de aansluiting van de toets op het nascholingsprogramma (zie Bijlage 13). Tenslotte werd de deelnemers gevraagd een korte vragenlijst te beantwoorden bij de retentietoets. Deze werd apart ontwikkeld voor de groep die aan de nascholing had deelgenomen (huisartsen zowel als huisartsen-in-opleiding) en voor de groepen die alleen een toets hadden ingevuld (zie Bijlage 16). In de vragenlijst bij de retentietoets werden vragen opgenomen over eventuele veranderingen in het handelen wat betreft diagnostiek of beleid op het gebied van hypercholesterolemie, en over het oordeel van de deelnemers achteraf over de nascholing (alleen voor de nascholingsgroepen). Tevens werd gevraagd of men in de periode tussen de nascholing, c.q. eerste toetsing, en het moment waarop de retentietoets werd afgenomen, op enigerlei wijze extra aandacht had geschonken aan het onderwerp cholesterol. Alle vragenlijsten werden voorgecodeerd ten behoeve van een anonieme gegevensverwerking.

Het aantal *deelnemers* lag gedeeltelijk vast, omdat werd aangesloten bij het implementatie-onderzoek naar de NHG-Standaard Cholesterol. Een groep van 16 huisartsen nam in het kader van deze gekoppelde onderzoeksopzet deel aan de nascholing, inclusief voor- en natoets, en aan de daaropvolgende verdere interventie en de retentie-kennistoets. De controlegroep uit het implementatie-onderzoek bestond uit 14 huisartsen, en maakte, als vergelijkingsgroep in het kennistoetsingsonderzoek, alleen de voortoets en de retentie-toets. Indeling van huisartsen in de implementatiestudie aan de interventiegroep en de controlegroep gebeurde volgens een matchingsprocedure, waarbij gelijkheid van beide groepen werd nagestreefd op de volgende kenmerken: al dan niet geautomatiseerd zijn van de praktijk, en praktijkvorm (van der Weijden, 1992). Daarnaast werden enkele huisartsengroepen bereid gevonden deel te nemen aan de nascholing plus toetsing, danwel alleen als vergelijkingsgroep aan de voor- en retentietoets. Bij deze groepen vond toewijzing van deelnemers aan de nascholings- danwel de vergelijkingsgroep niet at random plaats. Aan de toetsing en nascholing over cholesterol namen twee groepen huisartsen-in-opleiding deel. Twee andere groepen van huisartsen-in-opleiding deden als vergelijkingsgroep alleen aan de toetsing mee. De groepen waren afkomstig van twee verschillende instituten en verkeerden in dezelfde opleidingsfasen. De huisartsen-in-opleiding hadden nog geen formeel onderwijs over cholesterol gehad.

Een overzicht van de aantallen deelnemers in de verschillende groepen is weergegeven in tabel 7.6.

Tabel 7.6. *Deelnemers kennistoetsing cholesterol.*

Deelnemers	Aantal		
	voortoets	natoets	retentietoets
<i>Nascholingsgroepen</i>			
huisartsen (implementatiestudie)	16	16	15
overige huisartsen	12	12	11
huisartsen-in-opleiding	20	20	18
totaal	48	48	44
<i>Vergelijkingsgroepen</i>			
huisartsen (controlegroep implementatiestudie)	14		14
overige huisartsen	20		17
huisartsen-in-opleiding	22		16
totaal	56		47
Totaal	104	48	91

Uiteindelijk hebben 104 huisartsen en huisartsen-in-opleiding de voortoets gemaakt. Van hen behoorden 48 deelnemers (28 huisartsen en 20 huisartsen-in-opleiding) tot de nascholingsgroepen. De retentietoets werd acht maanden later door 91 deelnemers ingevuld. Dat betekent een uitval van ongeveer 10%.

In tabel 7.7 zijn de belangrijkste achtergrondkenmerken van de deelnemers weergegeven.

Tabel 7.7. *Achtergrondkenmerken deelnemers kennistoetsing cholesterol.*

Achtergrondkenmerken	Nascholingsgroepen		Vergelijkingsgroepen	
	Huisartsen (N=28)	Huisartsen-in- opleiding (N=20)	Huisartsen (N=34)	Huisartsen-in- opleiding (N=22)
<i>Geslacht</i>				
man	81%	40%	73%	41%
vrouw	19%	60%	27%	59%
<i>Gemiddelde leeftijd</i>	43 jaar (std 7)	30 jaar (std 3)	40 jaar (std 5)	31 jaar (std 2)
<i>Praktijksituering</i>				
grote stad	—	10%	18%	32%
stad	56%	15%	38%	23%
versted. platteland	26%	30%	21%	27%
platteland	19%	45%	24%	18%
<i>Werkverband</i>				
solist	33%	60%	15%	64%
duo/associatie	19%	35%	44%	14%
groepspraktijk/gez.centrum	48%	5%	42%	23%

De huisartsen uit de nascholingsgroep zijn gemiddeld iets ouder dan de huisartsen uit de vergelijkingsgroep. De nascholings- en vergelijkingsgroep van huisartsen-in-opleiding wijken in dit opzicht niet van elkaar af. De nascholings- en vergelijkingsgroep van huisartsen zijn wat betreft praktijksituering redelijk vergelijkbaar. In de vergelijkingsgroep zijn naar verhouding meer huisartsen werkzaam in duo- of associatiepraktijken, terwijl in de nascholingsgroep meer huisartsen werkzaam zijn als solist of in gezondheidscentra. Wat betreft de huisartsen-in-opleiding zijn de nageschoolden vaker werkzaam op het platteland en in duo/associatiepraktijken. De vergelijkingsgroep is vaker werkzaam in de grote stad en in groepspraktijken of gezondheidscentra.

Bij de opzet van het *nascholingsprogramma* werd rekening gehouden met de criteria die de literatuur aanreikt voor de opzet van effectieve en efficiënte nascholing (Harden & Laidlaw, 1992). Dat wil zeggen dat de nascholing werd georganiseerd in kleine groepen, dichtbij de praktijken van de deelnemende huisartsen. Het onderwijs werd gegeven volgens een vaste formule in de vorm van een lezing gevolgd door casusbesprekingen. Door de keuze voor een lezing werd gegarandeerd dat bij elke groep dezelfde kennis aan bod kwam. Het statische karakter ervan werd enigszins doorbroken door ruim gelegenheid te bieden tot het stellen van vragen en discussie over het onderwerp. Aan het eind werd papieren casuïstiek aangeboden die men onderling moest oplossen. Het nascholingsprogramma kende een vaste systematiek met een logische opbouw: epidemiologische gegevens, case-finding, diagnostiek en tenslotte beleid ten aanzien van hypercholesterolemie. Vragen of onduidelijkheden die tijdens de nascholing ontstonden en/of niet konden worden opgelost, werden voorgelegd aan een van de samenstellers van de NHG-Standaard. Deze was daartoe op een tevoren afgesproken moment telefonisch bereikbaar. Met deze procedure werd beoogd de efficiëntie van de nascholing te bevorderen. Door de docent werd bijgehouden of alle elementen die in het programma aan bod moesten komen, ook daadwerkelijk aan de orde waren geweest. Het gehele nascholingsprogramma, inclusief toetsen, in te vullen vragenlijsten en pauze nam 3 uur in beslag.

De voortoets werd direct voorafgaand, de natoets meteen na afloop van het nascholingsprogramma ingevuld. Deze procedure verschilde in zoverre van die bij de toetsing rondom vaginaal bloedverlies, dat nu de voortoets ook werd ingevuld onder standaard-examencondities, dat wil zeggen groepsgewijs, onder toezicht en zonder naslagmateriaal. In het nascholingsprogramma was hiervoor speciaal tijd gereserveerd. De vergelijkingsgroepen maakten de kennistoets hetzij onder standaard-examencondities, hetzij individueel onder toezicht. In het laatste geval maakten onderzoeksmedewerkers (na een telefonische toelichting) een spreekuurspraak met de betrokken huisarts, die vervolgens ter plekke en in aanwezigheid van de onderzoeksmedewerker, de toetsformulieren invulde. De retentietoets werd acht maanden later gemaakt, eveneens onder examenomstandigheden.

Na afloop van de toetsing werd de kennistoets op grond van de beschikbare gegevens onderzocht op onduidelijke of inadequate vragen. Eén vraag is ten gevolge hiervan vervallen. De feedback en nadere analyses zijn gebaseerd op de geschoonde toetsversie van 29 vragen. Eerst werden de gemiddelde goed-min-foutscores en standaarddeviaties berekend van de verschillende deelnemersgroepen op de voor-, na- en retentietoets. Omdat de voor- en de natoets en de nascholing tijdens dezelfde bijeenkomst plaatsvonden, was er geen uitval tussen voor- en natoets. De retentietoets vond acht maanden later plaats. De invloed van eventuele uitvallers bij de retentietoets werd onderzocht door hun achtergrondkenmerken en hun gemiddelde scores op de voor- of natoets te vergelijken met de kenmerken en scores van

degenen die de retentietoets wel maakten. Uit deze vergelijking bleek niet dat de uitvallers op enigerlei wijze afweken van degenen die de retentietoets wel invulden. Ook de gemiddelde scores op voor- en natoets veranderden niet, of de uitvallers op de retentietoets daar wel of niet bij werden geteld.

Om inzicht te krijgen in de samenhang tussen scoreverloop en kenmerken van de deelnemersgroep, werden de volgende analyses uitgevoerd. Allereerst werd de correlatie berekend tussen de achtergrondvariabelen en de goed-min-foutscore op de voortoets (Pearson produkt-moment correlatiecoëfficiënten). De literatuur geeft aan dat voor het bepalen van factoren die van invloed zijn op de kennisontwikkeling, het hanteren van verschillcores (dat wil zeggen het verschil tussen eindscore en beginscore) problemen oplevert (Bereiter, 1963), en bovendien vaak niet nodig is. Aanbevolen wordt om de voortoetsscore als onafhankelijke predictorvariabele op te nemen in een regressie-analyse waarbij de eindscore de afhankelijke variabele is (Cronbach & Furby, 1970; Linn, 1988).

In een regressie-analyse werd daarom nagegaan wat de bijdrage was van verschillende kenmerken van de deelnemers, waaronder de voortoetsscore, op de goed-min-foutscore die werd behaald op de retentietoets over cholesterol. De volgende onafhankelijke variabelen werden, zo mogelijk gedichotomiseerd, in de regressievergelijking betrokken:

- de goed-min-foutscore op de voortoets,
- de leeftijd en het geslacht van de deelnemers,
- de deelname aan nascholings- of aan vergelijkingsgroep,
- fase in de carrière (huisarts-in-opleiding versus huisarts),
- de praktijkvorm (solist versus niet-solist),
- de geografische situering van de praktijk,
- zelf aangegeven verandering in handelen na deelname aan toetsing,
- zelf aangegeven extra aandacht aan cholesterol na toetsing.

Met behulp van multiële regressie-analyse werd nagegaan welke combinatie van onafhankelijke variabelen het meeste bijdroeg aan de verklaarde variantie van de afhankelijke variabele, de score op de retentietoets.

- resultaten constructvalidering

In tabel 7.8 zijn van de verschillende groepen deelnemers de resultaten op de cholesteroltoetsen weergegeven.

In tabel 7.8 zijn de gegevens opgenomen van alle deelnemers die aan de voor-, na-, resp. retentietoets hebben deelgenomen. Tussen de voor- en de natoets bestond geen uitval, omdat deze toetsen werden gemaakt als 'onderdeel' van de nascholing. De retentietoets liet enige uitval zien (N=13, zie tabel 7.6). Noch achtergrondkenmerken noch gemiddelde scores op de voortoets van deze uitvallers wezen erop dat zij in enig opzicht afweken van de overige deelnemers.

Het verschil in gemiddelde goed-min-foutscore op de voortoets tussen de totale groep huisartsen die aan de nascholing deelnam en de totale groep huisartsen-in-opleiding die aan de nascholing deelnam (45, resp. 33%) was statistisch significant ($p < 0.05$). De overige verschillen tussen de gemiddelde scores op de voortoets van de verschillende deelnemersgroepen waren statistisch niet significant.

Tabel 7.8. Gemiddelde goed-min-foutscores (plus standaarddeviaties) van ervaren huisartsen en huisartsen-in-opleiding op voortoets, natoets en retentietoets cholesterol.

Deelnemers	Gemiddelde goed-min-foutscore in percentages (met standaarddeviatie)		
	voortoets	natoets	retentietoets
<i>Nascholingsgroepen</i>			
huisartsen (implementatiestudie, N=16)	42 (20)	73 (13)	51 (14)
overige huisartsen (N=12)	48 (18)	72 (17)	54 (14)
totaal huisartsen (N=28)	45 (19)	73 (15)	53 (14)
huisartsen-in-opleiding (N=20)	33 (15)	73 (11)	49 (18)
<i>Vergelijkingsgroepen</i>			
huisartsen (controlegroep implementatiestudie, N=14)	49 (16)		53 (14)
overige huisartsen (N=20)	41 (18)		41 (16)
totaal huisartsen (N=34)	44 (17)		47 (16)
huisartsen-in-opleiding (N=22)	38 (16)		39 (22)

Op de natoets waren de gemiddelde goed-min-foutscores van de huisartsen en de huisartsen-in-opleiding die de nascholing volgden, gelijk. Dat betekent dat bij de huisartsen-in-opleiding een grotere kennistoename werd gemeten dan bij de ervaren huisartsen. Het verschil bedroeg gemiddeld bijna 30% voor de huisartsen en 40% voor de huisartsen-in-opleiding. De gemiddelde scores op de retentietoets zijn voor de nascholingsgroepen weer lager dan op de natoets, maar de scores zijn wel hoger dan de voortoetsscores.

Opvallend is het verschil in scoreverloop tussen de groepen huisartsen en huisartsen-in-opleiding. De huisartsen houden gemiddeld na acht maanden ongeveer eenderde deel van de toegenomen score na de nascholing vast, de huisartsen-in-opleiding houden gemiddeld bijna de helft van de toegenomen score vast. De gemiddelde scores van de verschillende vergelijkingsgroepen zijn niet wezenlijk veranderd. Overigens is alleen het verschil in gemiddelde voortoets- en retentietoetsscore van de huisartsen-in-opleiding die deelnamen aan de nascholing statistisch significant ($p < 0.05$). De gemiddelde score op de retentietoets is voor de groep huisartsen wel hoger dan de gemiddelde score op de voortoets, maar het verschil is niet significant. Ook voor de groep huisartsen uit de implementatiestudie, die een intensieve praktijkbegeleiding kreeg in de periode na de nascholing, was de score op de retentietoets niet significant hoger dan de beginscore. Deze groep vertoonde echter wel een grotere stijging van de gemiddelde score op de voortoets en de retentietoets, dan de andere huisartsen.

Evenals bij de toetsing over vaginaal bloedverlies is het oordeel van de deelnemers over de aansluiting tussen de toets en het nascholingsprogramma ingewonnen. In tabel 7.9 is het oordeel van de deelnemers aan de nascholing over cholesterol weergegeven.

Tabel 7.9. *Oordeel nascholingsdeelnemers over aansluiten kennistoets cholesterol op nascholingsprogramma.*

Opvattingen	Waardering (in percentages)		
	(geheel) eens	neutraal/geen mening	(geheel) oneens
<i>Inhoud voortoets is goede afspiegeling nascholingsprogramma</i>			
huisartsen (N=28)	93	7	—
huisartsen-in-opleiding (N=20)	80	10	10
totaal (N=48)	88	8	4
<i>Inhoud natoets is goede afspiegeling nascholingsprogramma</i>			
huisartsen (N=28)	89	11	—
huisartsen-in-opleiding (N=20)	74	16	11
totaal (N=48)	83	13	4

Een ruime meerderheid van zowel de huisartsen als de huisartsen-in-opleiding vond dat de toetsen een goede representatie vormden van de inhoud van de nascholing. Evenmin als bij de toetsen over vaginaal bloedverlies, viel het niet alle deelnemers aan de cholesteroltoetsing op dat voor- en natoets inhoudelijk identieke toetsen waren.

Tabel 7.10. *Zelfrapportage ten aanzien van verandering in handelen (opsporings- en behandelingsbeleid) na de nascholing over cholesterol.*

Deelnemers nascholing cholesterol (N=42)	Verandering in handelen wat betreft cholesterolproblematiek (antwoorden in percentages)		
	(geheel) eens	neutraal/ geen mening	(geheel) oneens
<i>Actievere opsporing hypercholesterolemie</i>	33	22	45
<i>Ander beleid bij patiënten met hypercholesterolemie</i>	50	22	28

In tabel 7.10 worden de antwoorden van de deelnemers aan de retentietoets weergegeven op de vragen over eventuele veranderingen in hun beleid ten aanzien van hypercholesterolemie in de periode tussen de nascholing en de retentietoets.

Eenderde van de nascholingsdeelnemers gaf aan een actiever opsporingsbeleid te zijn gaan volgen en de helft van de deelnemers gaf aan na de nascholing een ander beleid te zijn gaan volgen ten aanzien van patiënten met hypercholesterolemie.

Driekwart van de deelnemers aan de nascholing gaf aan achteraf een positief oordeel te hebben over het nascholingsprogramma. Op de vraag of ze in de periode na de nascholing extra aandacht aan het thema cholesterol hadden geschonken, antwoordde 42% van de deelnemers bevestigend. De extra aandacht werd op diverse wijze vorm gegeven, bijvoorbeeld

in gesprekken over het thema met huisartsen-in-opleiding, of in het nalezen van de NHG-Standaard of andere literatuur.

De mate waarin de verschillende variabelen gezamenlijk bijdroegen aan de voorspelling van de retentietoetsscore, is onderzocht met behulp van een stapsgewijze multiële regressie-analyse. Conform de aanbevelingen uit de literatuur is de score op de voortoets in de analyse betrokken als onafhankelijke predictorvariabele (Cronbach & Furby, 1970). Dat betekent eigenlijk dat in de analyse rekening wordt gecorrigeerd voor de invloed die de score op de voortoets uitoefent op de kennisgroei.

Het resultaat van de stapsgewijze multiële regressie-analyse wordt weergegeven in tabel 7.11. Alleen de variabelen die een significante bijdrage hadden, zijn in het model opgenomen.

Tabel 7.11. Resultaat stapsgewijze multiële regressie-vergelijking met gemiddelde goed-min-foutscore op de retentietoets als afhankelijke variabele ($N=104$).

Onafhankelijke variabelen	proportie verklaarde variantie
Goed-min-foutscore op voortoets	0.11
Deelname aan nascholing	0.07
Totaal	0.18

Alleen deelname aan de nascholing levert in combinatie met de goed-min-foutscore op de voortoets, een statistisch significante bijdrage aan de verklaarde variantie op de retentietoets. De twee variabelen verklaren gezamenlijk 18% van de variantie. Geen enkele andere variabele bereikte het significantieniveau van $p < 0.15$ dat gold om tot het model te worden toegelaten. Dat wil zeggen dat 82% van de variantie door andere factoren wordt verklaard.

7.4. CONCLUSIES VALIDITEIT ONDERWERPGEBONDEN KENNISTOETSEN

De procedure waarvoor bij de samenstelling van de onderwerpgebonden toetsen is gekozen, vormt een garantie dat de toetsen goed zijn afgestemd op de relevante elementen uit de NHG-Standaarden. Door deze adequate dekking van de stof, wordt de inhoudsvaliditeit van de toetsen ondersteund. Enig voorbehoud geldt voor de toets over vaginaal bloedverlies, omdat de omzetting van onderwijsdoelen in 'key features', waarop de toetsvragen zijn gebaseerd, is uitgevoerd door slechts één deskundige. Voor de inhoudsvaliditeit van de kennistoets over cholesterol stond de gekozen methode van constructie, waarbij verschillende huisartsen-experts de relevante 'key features' bepaalden, borg. De vragen voldeden aan de eisen die worden gehanteerd door het Samenwerkingsverband.

De toetsen die zijn gebruikt als voormeting, nameting en (bij het thema 'cholesterol') retentiemeting, waren identiek en omvatten alle kennisaspecten die door de experts relevant waren genoemd. Hiermee werd een meettechnisch probleem ten gevolge van niet equivalente toetsen voorkomen. Dit kan bijvoorbeeld tot uiting komen in een verschil in moeilijkheidsgraad tussen de toetsen, of in inhoudelijke verschillen, hetgeen de toepasbaarheid van de toetsen om kennisverschillen aan te tonen, bemoeilijkt (Linn, 1988). Gebruik van identieke toetsen heeft echter ook nadelen: de toetsen kunnen worden 'herkend' (Weiss, 1972). Dat een

dergelijk effect het onderhavige onderzoek problemen heeft opgeleverd, lijkt onwaarschijnlijk. Een eventueel leereffect (van de voortoets) werd in het kader van de nascholing niet als versturende variabele beschouwd. Het was immers de bedoeling dat kennisoverdracht zou plaatsvinden. De antwoordsleutel was bovendien op het moment van de tweede toetsing niet bekend. In het gekozen design blijft de vraag onopgelost in hoeverre de stijging in score tussen voor- en natoets geheel kan worden toegeschreven aan het gevolgd hebben van het onderwijsprogramma of gedeeltelijk moet worden toegeschreven aan een gecombineerd effect van toetsing en scholing. Dat een eventueel leereffect van de voortoets invloed had op de scores op de retentietoets, is onwaarschijnlijk gezien de gelijkblijvende gemiddelde scores van de vergelijkingsgroepen.

De relatief grote spreiding van de gemiddelde scores op beide toetsen, zowel voor de voortoetsen, de natoetsen, als de retentietoets, geeft aan dat de toetsen niet te kampen hebben met een zogenaamd 'vloer-' of 'plafond'effect (Linn, 1988). In het eerste geval zijn de toetsen te moeilijk om als instrument te kunnen dienen voor het bepalen van een beginniveau. In het tweede geval zijn de toetsen te gemakkelijk om zinvol te kunnen fungeren als instrument om effecten te meten: de scores clusteren dan rond de maximaal haalbare score, zodat de toets de deelnemers niet de gelegenheid geeft om hun werkelijke groei in kennis te laten zien.

Voor zover kon worden vastgesteld dat het onderwijs adequaat was, laten de resultaten van de kennistoetsing rondom de nascholing zowel voor het thema vaginaal bloedverlies, als voor cholesterol zien dat de nascholing een direct effect op de kennis over het thema heeft. De retentietoets bij het thema cholesterol maakt duidelijk dat dit effect ook weer wegebt. Hoewel bij alle groepen die deelnamen aan de nascholing, na acht maanden nog een hogere gemiddelde score dan de gemiddelde initiële score werd aangetroffen, was dit effect alleen bij de groep huisartsen-in-opleiding die deelnamen aan de nascholing, statistisch significant ($p < 0.05$). De bevinding kwam daarmee overeen met de verwachtingen en ondersteunt de constructvaliditeit van de toets. Deze wordt ook ondersteund door de bevinding bij de toetsing over cholesterol, dat de praktiserende huisartsen op de voortoets gemiddeld statistisch significant hogere scores behaalden dan de huisartsen-in-opleiding. De NHG-Standaard over cholesterol, waar de toets op was gebaseerd, was immers al geruime tijd gepubliceerd. Verwacht mocht worden dat de zittende huisartsen meer kans hadden gehad zich te verdiepen in de inhoud van deze NHG-Standaard, dan de huisartsen-in-opleiding. Deze laatsten hadden, op het moment van de toetsing, nog geen formeel onderwijs over hypercholesterolemie gehad. De afname in gemiddelde score op een kennistoets enige tijd na afloop van de nascholing is geen verrassende bevinding. Recent onderzoek in Nederland bij nascholingsgroepen van huisartsen wees op hetzelfde fenomeen (Kempen e.a., 1992; Axt-Adam e.a., 1993; Soeters e.a., 1993). Deze laatste onderzoeken meldden zelfs het geheel achterwege blijven van enige kennistoename.

Eén groep huisartsen die de nascholing over cholesterol volgde, werd daarna in de praktijk intensief begeleid ten aanzien van het handelen met betrekking tot hypercholesterolemie. Tegen de verwachtingen in wezen de gemiddelde scores van deze groep op de retentietoets over cholesterol, er nauwelijks op dat deze intensieve begeleiding tot hogere scores leidde op de kennistoets. Mogelijk hangt deze bevinding samen met de onderwerpkeuze, met de termijn van ongeveer drie maanden tussen de laatste praktijkbegeleiding en de retentietoetsafname, of met andere factoren.

De regressie-analyse die voor de constructvalideringsstudie naar cholesterol werd gemaakt, maakt duidelijk welke invloed verschillende variabelen hebben op de kennis die acht maanden na afloop van de nascholing resteert (Linn, 1988). Gecorrigeerd voor het effect van de voortoetscore op de groei in kennis, bleek alleen een statistisch significant effect te bestaan voor het al dan niet daadwerkelijk gevolgd hebben van het nascholingsprogramma. Andere variabelen zoals leeftijd, geslacht, praktijkvorm of -situering, of de eigen inschatting van een veranderd handelen ten aanzien van hypercholesterolemiepatiënten, hadden geen significant effect op de kennisontwikkeling.

Wat betreft de omvang van de invloed van de factoren die bijdragen aan de verklaring van effecten van de nascholing, blijkt uit een stapsgewijze multiële regressie-analyse, dat de goed-min-foutscore op de voortoets en het hebben deelgenomen aan de nascholing, tezamen 18% van de verklaarde variantie voor hun rekening nemen. Geen andere variabelen voegen daar nog significant iets aan toe. Dat betekent dat het grootste deel van de variantie op de retentietoets moet worden verklaard uit andere dan achtergrondkenmerken van de deelnemers aan de nascholing.

De laatste conclusie betreft de invloed van de instructiewijze op de onderwijs- of nascholingseffectiviteit. De literatuur geeft aan dat behalve de intensiteit (de hoeveelheid) van het onderwijs, ook de kwaliteit van de instructie en van de docenten een belangrijke rol speelt bij de effectiviteit van nascholingsprogramma's (Cauffman et al, 1985; Beaudry, 1989). Observatie van de nascholing bij het thema vaginaal bloedverlies, maakte duidelijk dat in één van de groepen de informatie-overdracht gebrekkig verliep. De invloed daarvan op de effectiviteit van de nascholing kon in het onderzoek worden betrokken. De resultaten bevestigen het belang van een goede instructie als voorwaarde voor effectiviteit. Over het deel van het onderwijs dat aantoonbaar niet of slecht tijdens de nascholing aan bod was gekomen, werden geen kennisverschillen gemeten. Hieruit kan worden afgeleid hoe belangrijk het is dat onderwijs en toets inhoudelijk goed op elkaar zijn afgestemd, en tevens dat het van groot belang is goed te letten op wat er precies in de nascholing gebeurt.

Hoofdstuk 8

BETROUWBAARHEID ONDERWERPGEBONDEN KENNISTOETSEN

SAMENVATTING

De betrouwbaarheid van de onderwerpgebonden kennistoetsen is onderzocht met behulp van analyses die zijn gebaseerd op de generaliseerbaarheidstheorie. De betrouwbaarheid van individuele goed-min-foutscores en van gemiddelde groepsscores voor de onderwerpgebonden toetsen is op dezelfde wijze bepaald als de betrouwbaarheid van de scores op de algemene kennistoets.

De betrouwbaarheid van individuele toetsscores op de exemplarische onderwerpgebonden toetsen bleek vanuit norm- of relatief georiënteerd perspectief op de toets over vaginaal bloedverlies .53 (standaardmeetfout (SEM) 16), en op de toets over cholesterol .41 (SEM 15). Vanuit domein- of absoluut georiënteerd perspectief bedroeg deze .49 (SEM 18), resp. .38 (SEM 15). Deze bevindingen kwamen overeen met de schattingen die in hoofdstuk 3 zijn gegeven op basis van de gegevens uit de algemene kennistoets. De betrouwbaarheid van de gemiddelde groepsscores op de onderwerpgebonden kennistoetsen leverde een SEM op van 4, resp 8% voor het relatieve en het absolute betrouwbaarheids perspectief voor de kennistoets over vaginaal bloedverlies, en een SEM van 3, resp. 6% vanuit het relatieve en vanuit het absolute perspectief voor de kennistoets over cholesterol.

Geconcludeerd wordt dat de onderwerpgebonden kennistoetsen onvoldoende betrouwbaar zijn om er consequenties voor individuele huisartsen aan te mogen verbinden. De gemiddelde groepsscores zijn betrouwbaar genoeg om de toetsen te kunnen toepassen als middel om de effectiviteit van nascholing in termen van kennisonwikkeling mee te onderzoeken.

8.1. INLEIDING

Informatie over de betrouwbaarheid van onderwerpgebonden kennistoetsen is vanuit een aantal verschillende perspectieven van belang. Bij gebruik van de toets als diagnostisch hulpmiddel om sterke en zwakke kanten van individuen op te sporen, geeft informatie over de betrouwbaarheid van individuele scores immers inzicht in de mate waarin de scores een consistente en nauwkeurige weergave vormen van de kennis over het onderwerp waarop de toets betrekking heeft. Betrouwbaarheid van groepsgemiddelden is van belang om uitspraken te kunnen doen over de mate waarin de gemiddelde groepsscores informatie bieden over sterke en zwakke kanten op groepsniveau. Omdat de onderwerpgebonden toetsen vooral tot functie hebben de effectiviteit in termen van kennistoename van het volgen van een nascholingsprogramma zichtbaar te maken, is vooral de betrouwbaarheid van belang in termen

van consistentie en nauwkeurigheid van de gemiddelde groepsscores (Feldt & Brennan, 1989). Bij voldoende betrouwbaarheid van de groepsgemiddelden mag de toets in dit laatste geval worden gebruikt als programma-evaluatiemiddel (Hambleton, 1988). Evenals in de studie naar de betrouwbaarheid van de algemene huisartsgeneeskundige kennistoets, is ervoor gekozen om de betrouwbaarheidsschattingen voor deze verschillende perspectieven te baseren op de generaliseerbaarheidstheorie (Cronbach et al, 1972; Brennan, 1983; Allal, 1988).

In de volgende paragrafen wordt de opzet en het resultaat beschreven van het onderzoek naar de betrouwbaarheid van de onderwerpgebonden kennistoetsen.

Onderzoeksvraag III.2.:

Wat is de betrouwbaarheid van onderwerpgebonden kennistoetsen die bij specifieke nascholingsprogramma's worden ontwikkeld?

Opzet betrouwbaarheidsstudie:

- generaliseerbaarheidsanalyse voor individuele toetsscores
- generaliseerbaarheidsanalyse voor gemiddelde groepsscores

8.2. ONDERZOEKSOPZET BETROUWBAARHEIDSSTUDIE ONDERWERPGEBONDEN KENNISTOETSEN

Voor de exemplarisch ontwikkelde onderwerpgebonden kennistoetsen over vaginaal bloedverlies en cholesterol zijn betrouwbaarheidsstudies uitgevoerd op basis van de generaliseerbaarheidstheorie. Naast betrouwbaarheidscoëfficiënten ter bepaling van de reproduceerbaarheid van de scores, werden standaardmeetfouten (SEM's) geschat ter bepaling van de nauwkeurigheid van de scores. In de analyses werden zowel de betrouwbaarheid van de individuele scores als de betrouwbaarheid van de groepsgemiddelden nagegaan.

Het *design* voor de generaliseerbaarheidsstudie op individueel niveau, was een 'one facet design' (Cronbach et al, 1972; Brennan, 1983). De volgende variantiebronnen werden in het design opgenomen: 'personen', 'items' en de interactie tussen personen en items, die in dit design niet kan worden onderscheiden van de overige error. 'Personen' vormen het object van meting; 'items' vormen het 'facet' ofwel de conditie waaronder wordt gemeten (zie Bijlage 3 voor een toelichting op de terminologie).

Variantiebronnen

P (personen)

I (items)

PI (interactie tussen personen en items)

Toelichting:

- personen (P): verschillen tussen personen vormen het object van meting. Idealiter zou deze factor de belangrijkste bijdrage moeten leveren aan de totale toetsbetrouwbaarheid;
- items (I): hiermee worden de verschillen in moeilijkheidsgraad tussen de items aangeduid;

- personen en items (PI): het interactie-effect tussen personen en items geeft aan in welke mate personen anders gerangordend worden door de verschillende items. Dit effect is in dit design niet te onderscheiden van allerlei andere foutenbronnen die de score beïnvloeden.

Voor de onderwerpgebonden toetsen was op voorhand geen logische groepsindeling van huisartsen beschikbaar. Daarom is voor de betrouwbaarheid van de gemiddelde groepsscores, conform de methode die is toegepast bij de betrouwbaarheidsbepaling van groepsgemiddelden voor de algemene kennistoets, de standaardmeetfout (SEM) als indicator gehanteerd (zie hoofdstuk 3). De SEM voor groepsgemiddelden is berekend op basis van de SEM voor individuen (Brennan, 1983).

De betrouwbaarheidsanalyses zijn uitgevoerd op de gegevens van de toetsen over 'vaginaal bloedverlies' en 'cholesterol'. De betrouwbaarheidsanalyses werden uitgevoerd voor de scores op de geschoonde toetsversies. De geschoonde toets over vaginaal bloedverlies bestond uit 25 vragen, en de geschoonde toets over cholesterol bestond uit 29 vragen.

De vraagvorm in de onderwerpgebonden toetsen was dezelfde als in de algemene huisartsgeneeskundige kennistoets werd toegepast. De scores werden op identieke wijze bepaald als de scores op de algemene toets. De standaardmeetfouten werden, conform de procedures die werden toegepast bij de algemene huisartsgeneeskundige kennistoets (zie hoofdstuk 3), uitgedrukt in dezelfde schaal als de scores, dat wil zeggen in percentages. Een praktisch voordeel van het hanteren van dezelfde vraagvorm en dezelfde scoreberekening was dat gebruik gemaakt kon worden van dezelfde gegevensverwerkings- en analyseprogramma's.

De betrouwbaarheidsberekeningen zijn gebaseerd op de gegevens van de toetsen die voorafgaand aan de nascholing werden afgenomen. Om de betrouwbaarheidsschattingen te optimaliseren, werden de gegevens van alle huisartsen die hetzij als onderwijsgroep, hetzij als vergelijkingsgroep (zie hoofdstuk 7) aan deze toetsing deelnamen, in de berekeningen betrokken. Dit betekende dat de gegevens voor de berekening van de betrouwbaarheid op individueel niveau voor de toets over vaginaal bloedverlies afkomstig zijn van 42 huisartsen. Voor de individuele betrouwbaarheidsberekeningen van de toets over cholesterol waren gegevens van 62 huisartsen beschikbaar. Als groepsgrootte voor de berekening van de betrouwbaarheid van gemiddelde groepsscores werd een groep van 30 huisartsen genomen. Deze grootte stemt ongeveer overeen met de daadwerkelijke grootte van de groepen die waren betrokken bij het onderzoek naar de constructvaliditeit van de onderwerpgebonden toetsen (zie hoofdstuk 7).

8.3. RESULTATEN

8.3.1. Betrouwbaarheid individuele scores onderwerpgebonden kennistoetsen

De gemiddelde goed-min-foutscore van de huisartsen op de toets over vaginaal bloedverlies was 31% en op de toets over cholesterol 45% (zie hoofdstuk 7).

De geschatte variantiecomponenten van de individuele goed-min-foutscores van de huisartsen op de onderwerpgebonden kennistoetsen wordt weergegeven in tabel 8.1.

Tabel 8.1. *Geschatte variantiecomponenten en percentage van de totale variantie van individuele toetsscores van huisartsen op onderwerpgebonden toetsen.*

Variantiebron	Geschatte variantiecomponenten	Percentage van de totale variantie
<i>Kennistoets 'vaginaal bloedverlies'</i>		
P (N=42)	.03052	3.7
I (N=25)	.13009	15.7
PI	.66539	80.6
<i>Kennistoets 'cholesterol'</i>		
P (N=62)	.01459	2.1
I (N=29)	.07988	11.3
PI	.60985	86.6

Met de schattingen van de variantiecomponenten kon worden bepaald dat de betrouwbaarheid van de individuele toetsresultaten voor de kennistoets over vaginaal bloedverlies vanuit normgeoriënteerd (relatief) perspectief .53 bedroeg, en voor de toets over cholesterol .41. De standaardmeetfout (SEM) bedroeg 16, resp. 15. Vanuit domeingeoriënteerd (absoluut) perspectief bedroegen de betrouwbaarheidscoëfficiënten voor de kennistoets over vaginaal bloedverlies .49 (SEM 18), en voor de kennistoets over cholesterol .38 (SEM 15). Dat betekent dat voor de interpretatie van een individuele score van bijvoorbeeld 45 op de kennistoets over cholesterol (de gemiddelde score van de huisartsen) met een marge van 45 plus of min 2×15 rekening gehouden moet worden om het 95%-betrouwbaarheidsinterval van deze score te kunnen bepalen. Met andere woorden: voor een individuele score van 45 geldt een gebied dat ligt tussen 15 en 75 waarin met een zekerheid van 95% de ware score kan liggen.

8.3.2. Betrouwbaarheid gemiddelde groepsscores onderwerpgebonden kennistoetsen

Op basis van de resultaten van de betrouwbaarheidsschattingen van de individuele scores, werd de standaardmeetfout van gemiddelde groepsscores bepaald voor de toetsen over vaginaal bloedverlies en over cholesterol. Gekozen werd voor het berekenen van de betrouwbaarheid voor een groep van 30 personen. Het resultaat wordt weergegeven in tabel 8.2.

De standaardmeetfout is vanuit relatief (rangordenings-)perspectief 4% voor de toets over vaginaal bloedverlies, en 3% voor de toets over cholesterol. Vanuit absoluut (beheersings-) perspectief zijn deze getallen 8% voor de toets over vaginaal bloedverlies, en 6% voor de toets over cholesterol. Dat betekent bijvoorbeeld voor de toets over cholesterol, waarbij de gemiddelde score 45 bedroeg, dat de feitelijke score, voor een groep van 30 huisartsen, tussen 42 en 48 ligt (67%-betrouwbaarheidsinterval) of tussen 39 en 51 (95%-betrouwbaarheidsinterval).

Tabel 8.2. *Geschatte betrouwbaarheid, uitgedrukt in standaardmeetfouten (SEM), van gemiddelde groepsscores van deelnemers aan exemplarische onderwerpgebonden kennistoetsen.*

Onderwerpgebonden kennistoetsen	Standaardmeetfout (SEM) gemiddelde groepsscores (%)	
	Relatief perspectief	Absoluut perspectief
Kennistoets 'vaginaal bloedverlies' (25 vragen; groepsgrootte 30 huisartsen)	4	8
Kennistoets 'cholesterol' (29 vragen; groepsgrootte 30 huisartsen)	3	6

8.4. CONCLUSIES BETROUWBAARHEID ONDERWERPGEBONDEN KENNISTOETSEN

De standaardmeetfout (SEM) waarmee rekening moet worden gehouden bij de interpretatie van de individuele scores die op de onderwerpgebonden toetsen werden behaald, bedroeg ongeveer 15. De foutenmarge van individuele toetsscores is dus aanzienlijk: met marges van plus of min 30 ($2 \times \text{SEM}$) scorepunten moet worden gerekend om het gebied te traceren waarin, gegeven een bepaalde score van een individuele huisarts, diens ware score kan liggen. Deze marges zijn te ruim om aan de individuele scores conclusies te kunnen verbinden. De betrouwbaarheid van de exemplarisch ontwikkelde onderwerpgebonden kennistoetsen vanuit individueel perspectief is dan ook niet voldoende.

De betrouwbaarheidsgegevens zijn vergeleken met de gegevens die de schatting van de betrouwbaarheid opleverde op de algemene kennistoets (zie hoofdstuk 3). Met behulp van de generaliseerbaarheidstheorie is op basis van deze laatste gegevens geschat wat de betrouwbaarheid zou zijn van een relatief korte toets (Düsmen, 1992). De betrouwbaarheid werd geschat vanuit relatief en vanuit absoluut perspectief. De generaliseerbaarheidscoëfficiënt (relatief perspectief) werd geschat op .32 voor een toets van 30 vragen (SEM 14); de phi-coëfficiënt (absoluut perspectief) voor een toets van deze omvang werd geschat op .27 (SEM 15). De empirisch gevonden waarden voor de onderwerpgebonden kennistoetsen blijken wel iets hoger, maar wijken niet wezenlijk af. Voor de toets over vaginaal bloedverlies bedroeg de relatieve betrouwbaarheidscoëfficiënt .53 en de absolute betrouwbaarheidscoëfficiënt .49. Voor de toets over cholesterol bedroegen deze waarden resp. .41 en .38. De SEM was vrijwel identiek aan de op basis van de resultaten op de algemene toets geschatte waarde. Deze bevindingen bevestigen de schatting van de betrouwbaarheid van relatief korte kennistoetsen die werd gebaseerd op de gegevens van de algemene huisartsgeneeskundige kennistoets (zie hoofdstuk 3). Het betekent waarschijnlijk dat huisartsen in hun kennis op deelonderwerpen evenveel van elkaar verschillen als in hun algemene kennis.

Om de bruikbaarheid van de onderwerpgebonden toetsen vast te stellen als instrument om onderwijseffecten te bepalen, werd de betrouwbaarheid geschat van de gemiddelde groepsscore. Deze werd uitgedrukt in een standaardmeetfout, die werd gebaseerd op de standaardmeetfout van de individuele scores. Voor de toets over vaginaal bloedverlies werd

een SEM voor groepsgemiddelden vanuit het relatieve perspectief, gevonden van 4%; voor de toets over cholesterol was de SEM voor groepsgemiddelden vanuit dit perspectief 3%. Deze marges werden acceptabel geacht om de toetsen toe te passen als instrument om, op groepsniveau, onderwijseffecten mee te bepalen.

Bij deze resultaten kunnen enkele kanttekeningen worden gemaakt. De kennisaspecten die in de onderwerpgebonden toetsen aan de orde worden gesteld, zijn beschouwd als een random steekproef uit een groter domein. Mogelijk is deze aanname niet terecht, en omvatten de items tezamen het domein zelf. Ze vallen dan samen met alle kennis, die door experts op het betrokken gebied wordt beschouwd als datgene wat een praktizerend huisarts aan kennis paraat moet hebben. De items zouden dan moeten worden opgevat als 'fixed' variabelen (Suen, 1990). De betrouwbaarheidsschattingen hebben dan niet betrekking op een 'onbegrensd' domein van items, maar op een 'fixed' domein (Kane & Brennan, 1977). De casuïstische vraagvorm, waarin voor de praktijk relevante kennisaspecten aan de orde worden gesteld, levert argumenten die tegen deze redenering pleiten. Het is immers niet goed denkbaar dat alle relevante kennis op een bepaald gebied te vangen is in een beperkt aantal items, die voor alle huisartsen een beroep doen op dezelfde (ervarings)kennis. Bovendien staat de generaliseerbaarheidstheorie niet toe dat betrouwbaarheidsindices voor individuele toetsscores kunnen worden berekend vanuit een zogenaamd 'mixed' model waarin de items als 'fixed' variabelen worden opgevat (Brennan, 1983; Suen, 1990). Daarom is gekozen voor een 'conservatieve' benadering waarbij de items zijn beschouwd als 'random' variabelen (zie ook Bijlage 3). Voor de indices die de betrouwbaarheidsstudie opleverde, betekent dit dat het de laagst mogelijke schattingen zijn van de betrouwbaarheid. De werkelijke betrouwbaarheid ligt mogelijk hoger.

Hoofdstuk 9

PRAKTISCHE TOEPASBAARHEID ONDERWERPGEBONDEN KENNISTOETSING

SAMENVATTING

De praktische toepasbaarheid van onderwerpgebonden kennistoetsing is onderzocht door een inventarisatie van het oordeel van praktiserende huisartsen over de waarde van het gebruik van dergelijke kennistoetsen bij nascholingsprogramma's. Daarnaast is, evenals bij de algemene huisartsgeneeskundige kennistoetsing, nagegaan wat de uitvoerbaarheid van de toetsing was door de tijd en menskracht te inventariseren die noodzakelijk was toetsen te construeren, door de huisartsen te vragen hoeveel tijd het invullen van een toets kostte, en door de uitvoerbaarheid na te gaan van de analyse en verwerking van de toetsgegevens en feedbacksamenstelling.

De beschikbaarheid van de NHG-Standaarden en met name de door de experts aangegeven essentiële kenniselementen van deze Standaarden, bleken een efficiënt hulpmiddel bij de vraagconstructie. De huisartsen beoordelen kennistoetsing bij nascholing vrijwel unaniem als een zinvolle activiteit, vooral als smaakmaker of 'warming up' voor de nascholing. Onderwerpgebonden kennistoetsen lijken voor de gebruikers veel minder geschikt als hulpmiddel bij de keuze van bepaalde nascholingsprogramma's. Wat betreft de verwerking en analyse van de toetsgegevens en het samenstellen van feedback geldt, evenals bij de algemene kennistoets, dat dit geen problemen oplevert, onder de voorwaarde dat adequate gegevensverwerkende apparatuur en programmatuur beschikbaar is.

9.1. INLEIDING

In dit hoofdstuk wordt de opinie van ervaren huisartsen beschreven over de zin en de bruikbaarheid van specifieke, onderwerpgebonden toetsen voor nascholingsdoelen. Daarnaast wordt ingegaan op de praktische uitvoerbaarheid van het samenstellen van onderwerpgebonden toetsen.

Onderzoeksvraag III.3.:

Wat is de praktische toepasbaarheid in termen van acceptabiliteit en uitvoerbaarheid, van voor de nascholing bestemde, onderwerpgebonden kennistoetsen?

Opzet bepaling praktische toepasbaarheid:

- bepaling satisfactie huisartsen
- uitvoerbaarheid toetsconstructie en -verwerkingsmethodes

9.2. BEPALING PRAKTISCHE TOEPASBAARHEID ONDERWERPGEBONDEN KENNISTOETSING

Het oordeel van de deelnemers over de relevantie en de toepasbaarheid van de onderwerpgebonden kennistoetsen is ingewonnen met behulp van een vragenlijst met stellingen. Deze lijst werd ingevuld na afloop van het invullen van de natoets. De deelnemers konden op vijfpuntschalen hun oordeel geven over de inhoudelijke kwaliteit van de toetsen, de moeite die ze hadden met de vraagvorm en het aantal minuten dat het invullen van de toets(en) in beslag had genomen. Tenslotte werd hen gevraagd naar hun oordeel over de bruikbaarheid van onderwerpgebonden kennistoetsen in het kader van nascholing (zie Bijlage 13 voor de satisfactie-vragenlijst die bij de onderwerpgebonden toetsen is gehanteerd).

De vragen hadden betrekking op:

- de zinvolheid van de kennistoets als introductie op het nascholingsprogramma,
- de dekking van de toetsen (voor- resp. natoets) wat betreft de inhoud van het nascholingsprogramma,
- de moeilijkheid van de juist-onjuist-vraagtekenvorm van de toetsvragen,
- de bruikbaarheid van de onderwerpgebonden kennistoets voor het vaststellen van eigen lacunes op het betreffende gebied,
- het nut van kennistoetsen als 'warming up' voor een nascholingsprogramma,
- het nut van kennistoetsen als hulpmiddel om te kiezen voor deelname aan een bepaald programma,
- de algemene opinie van de huisartsen over de zinvolheid van kennistoetsing voorafgaand aan nascholingsactiviteiten.

De open vragen betroffen de benodigde invultijd voor de toets. Tevens kreeg men de gelegenheid opmerkingen te maken over de kennistoets en de toetsvragen.

De vragenlijst werd voorgecodeerd ten behoeve van anonieme verwerking.

De uitvoerbaarheid van de onderwerpgebonden kennistoetsing werd verder onderzocht door bij te houden hoeveel tijd en menskracht de verschillende toetsontwikkelingsfasen vergden. Eveneens werd de hoeveelheid tijd en menskracht bijgehouden die was gemoeid met de verzameling, verwerking en terugrapportage van de gegevens.

9.3. RESULTATEN

9.3.1. Acceptabiliteit onderwerpgebonden kennistoetsing

In tabel 9.1 zijn de opvattingen weergegeven van de deelnemende huisartsen aan de onderwerpgebonden toetsing rondom de nascholing 'vaginaal bloedverlies' en 'cholesterol'. De gegevens waren beschikbaar van 29 huisartsen die deelnamen aan het nascholingsprogramma over 'vaginaal bloedverlies' en van 28 huisartsen die deelnamen aan het nascholingsprogramma 'cholesterol'.

Tabel 9.1. *Opvattingen van ervaren huisartsen over onderwerpgebonden kennistoetsen. (kennistoets vaginaal bloedverlies: N=29, kennistoets cholesterol: N=28).*

Opvattingen	Waardering (in percentages)		
	(geheel) eens	neutraal/geen mening	(geheel) oneens
<i>Vraagvorm (JOV) is lastig</i> kennistoets 'vaginaal bloedverlies' kennistoets 'cholesterol'	– 11	17 18	83 71
<i>Toets is nuttig voor opsporen eigen kennislacunes</i> kennistoets 'vaginaal bloedverlies' kennistoets 'cholesterol'	78 86	17 7	7 7
<i>Toets is nuttig als 'warming up' voor nascholing</i> kennistoets 'vaginaal bloedverlies' kennistoets 'cholesterol'	97 92	3 4	– 4
<i>Toets is nuttig bij nascholingsprogrammakeuze</i> kennistoets 'vaginaal bloedverlies' kennistoets 'cholesterol'	48 55	45 37	7 7
<i>Kennistoetsing is niet zinvol</i> kennistoets 'vaginaal bloedverlies' kennistoets 'cholesterol'	– –	– 4	100 96

De opvattingen van de huisartsen over de acceptabiliteit en toepasbaarheid van de onderwerpgebonden kennistoets over vaginaal bloedverlies kwamen in grote lijnen overeen met de opvattingen van de deelnemers aan de kennistoets over cholesterol.

Beide groepen huisartsen beoordeelden het maken van een korte kennistoets voordat ze aan een nascholingsbijeenkomst deelnemen, als zinvol (100, resp. 96%). Een grote meerderheid (ongeveer 80%) beoordeelde de onderwerpgebonden kennistoetsen als een bruikbaar hulpmiddel om eigen kennislacunes vast te stellen en als goede 'warming up' voor het thema van de nascholing. Over de vraag of een kennistoets nuttig is als hulpmiddel om te kiezen voor deelname aan bepaalde nascholings thema's, werd genuanceerder gedacht. Slechts de helft van de huisartsen van wie gegevens beschikbaar waren, beantwoordde deze vraag positief.

9.3.2. Uitvoerbaarheid onderwerpgebonden kennistoetsing

Evenals bij de beschrijving van de praktische uitvoerbaarheid van de algemene huisartsgeneeskundige kennistoets worden hier drie invalshoeken gehanteerd ten aanzien van de uitvoerbaarheid van onderwerpgebonden kennistoetsen: de toetsconstructiefase, de afnamefase en de verwerkingsfase.

De toetsconstructiefase verliep voor beide onderwerpgebonden toetsen verschillend. Wel werd in beide gevallen ervoor gekozen dezelfde casuïstische juist-onjuist-vraagteken vraagvorm te gebruiken als in de algemene kennistoets. De productietijd van vragen voor de onderwerpge-

bonden toetsen kon in beginsel worden bekort omdat de inhoud van de respectievelijke NHG-Standaarden de literatuurreferentie vormde.

De voor de samenstelling van de kennistoets over cholesterol gevolgd consensusprocedure had tot gevolg dat met een termijn van ongeveer twee maanden rekening moest worden gehouden voordat de daadwerkelijke vraagproductie op gang kon komen. Het feit dat de consensusrondes leidden tot duidelijke kernbegrippen of key features waarop de toetsvragen inhoudelijk betrekking moesten hebben, betekende overigens weer wel dat deze vraagproductie zeer efficiënt en in relatief korte tijd kon plaatsvinden.

Met betrekking tot organisatorische en administratieve aspecten golden bij de samenstelling van de onderwerpgebonden kennistoetsen in beginsel dezelfde randvoorwaarden als bij de algemene kennistoets.

De tijd die voor de deelnemers aan de toetsen gemoeid was met het invullen van de toetsvragen, wordt voor beide toetsen weergegeven in tabel 9.2.

Tabel 9.2. Benodigde invultijd onderwerpgebonden kennistoetsen.

Deelnemers	Gemiddelde invultijd (in minuten)	
	voortoets	natoets
<i>Kennistoets 'vaginaal bloedverlies'</i> Nascholingsgroepen	14	11
<i>Kennistoets 'cholesterol'</i> Nascholingsgroepen	16	11

De benodigde invultijd voor de voortoets was ongeveer een kwartier. Opgemerkt moet worden dat de toets 'vaginaal bloedverlies' (26 vragen) iets korter was dan de kennistoets 'cholesterol' (30 vragen). De voortoets 'vaginaal bloedverlies' is door de deelnemers thuis, een of twee dagen voorafgaand aan de nascholing, ingevuld. De natoets werd meteen na afloop van de nascholing ingevuld. Voor de kennistoetsen over cholesterol gold dat beide toetsen op de nascholingsavond zelf werden ingevuld: de voortoets direct vóór en de natoets direct ná het programma. De vragen over de noodzakelijke invultijd werden voor beide toetsen meteen na afloop van de nascholing op het satisfactieformulier ingevuld.

De gegevens die zijn verzameld met de onderwerpgebonden kennistoetsen, zijn centraal verwerkt en geanalyseerd door het landelijk Samenwerkingsverband, waarbij gebruik is gemaakt van de beschikbare programmatuur en apparatuur voor toetsverwerking. Voor de feedback aan de huisartsen werd een speciaal formulier ontworpen waarop was aangegeven wat de individuele score was (zowel absoluut als procentueel) wat betreft juiste, onjuiste en met een vraagteken beantwoorde vragen, en, ter vergelijking, wat de groepsgemiddelden van deze scores waren voor de huisartsen die aan de toets hadden deelgenomen.

In het volgende schema worden de verschillende activiteiten in een tijdschema samengevat. In het schema zijn dezelfde elementen opgenomen als in het schema in hoofdstuk 4 voor de algemene huisartsgeneeskundige kennistoets.

Tabel 9.3. Overzicht activiteiten, menskracht en schatting van de tijdsinvestering bij onderwerpgebonden kennistoetsing (schatting voor toets van 30 vragen).

Activiteiten	Uitvoering	Tijdsinvestering
Formuleren toetsvragen <ul style="list-style-type: none"> - formuleren toetsvragen - literatuurverwijzing completeren - bespreken relevantie, vorm en inhoud - toetsvragen in standaard-lay out brengen 	inhoudsdeskundige nvt inhoudsdeskundigen administratie	5 dagdelen - 1,5 uur 2 dagdelen
Toetsconstructie <ul style="list-style-type: none"> - achterhalen key features - redigeren eindversie toets - toets drukklaar maken (- reproductie toets) 	inhoudsdeskundigen inhoudsdeskundige administratie (drukker)	PM 1,5 uur 1 uur (PM)
Toetsafname <ul style="list-style-type: none"> - toetsafname plannen - deelnemers werven - administratief bestand van deelnemers aanleggen - deelnemersinstructie formuleren - voorcoderen antwoordformulieren en vragenlijsten - individuele toetspakketten samenstellen en verzenden - antwoordsleutel samenstellen en verzenden 	inhoudsdeskundige administratie administratie inhoudsdeskundige administratie administratie administratie	PM PM per 100 pp: 1 dagdeel 1,5 uur per 100 pp: 2 dagdelen per 100 pp: 2 dagdelen per 100 pp: 1 dagdeel
Analyse en verwerking toetsgegevens <ul style="list-style-type: none"> - antwoordformulieren invoeren in optisch lezer - foutencorrectie uitvoeren, verwijderen items - (item)analyse uitvoeren - invoeren opmerkingen deelnemers - samenstellen definitieve toets voor feedback - invoeren satisfactiegegevens deelnemers - samenstellen overzichten satisfactie 	administratie methodoloog methodoloog administratie inhoudsdeskundigen administratie administratie	1,5 uur 1,5 uur 1,5 uur per 100 pp: 1 dagdeel 1,5 uur per 100 pp: 1 dagdeel PM
Feedbackproductie en -verzending <ul style="list-style-type: none"> - productie: individuele toetsresultaten - samenstellen groepsoverzichten: gemiddelde scores - verzenden feedback 	adm./methodoloog adm./methodoloog administratie	1,5 uur 1 uur per 100 pp: 1 dagdeel
Diversen <ul style="list-style-type: none"> - interpretatie, rapportage, advies 	inhoudsdesk./method.	PM

Dit overzicht beperkt zich tot een indruk van de tijdsinvestering die nodig is voor de feitelijke handelingen bij de toetsconstructie en -afname. Daarnaast moet rekening worden gehouden met benodigde tijd en menskracht voor de coördinatie van het proces en voor het uitvoeren van evaluatie-onderzoek naar de betekenis en uitkomsten van de onderwerpgebonden toetsing.

Het gegeven dat de literatuurreferentie in de vorm van de NHG-Standaard beschikbaar was, betekende een bekorting op de productietijd van de vragen. Gemiddeld vergde het vraagproductieproces ongeveer 40 minuten per vraag. Dit betekent een bekorting van gemiddeld een half uur per vraag (Tan & Pollemans, 1992).

9.4. CONCLUSIES PRAKTISCHE TOEPASBAARHEID ONDERWERPGEBONDEN KENNISTOETSING

De conclusie uit het onderzoek naar de praktische toepasbaarheid van onderwerpgebonden kennistoetsen voor nascholingsdoelen, is dat deze toetsen, vooral als wordt gekozen voor de methode zoals ontwikkeld bij de kennistoets 'cholesterol' binnen redelijk beperkte tijd kunnen worden samengesteld. Dat betekent wel dat een groep van 'experts' met betrekking tot het onderwerp beschikbaar moet zijn en bereid tot deelname aan de toetsontwikkeling. Voor de afname en verwerking geldt eigenlijk hetzelfde als voor de algemene huisartsgeneeskundige kennistoets. Als wordt beschikt over goedlopende procedures en over adequate programmatuur en apparatuur, bestaat er geen enkel probleem.

De huisartsen die deelnamen aan de onderwerpgebonden toetsen waren in meerderheid positief in hun oordeel over de toetsen, met name wat betreft hun gebruikswaarde als 'warming up' voor gerichte nascholing. Ook over de zinvolheid van kennistoetsing in zijn algemeenheid in de nascholing zijn de deelnemende huisartsen vrijwel unaniem positief. Over de rol van specifieke kennistoetsen bij het bepalen van de keuze voor een nascholingsprogramma zijn de deelnemers veel minder zeker. Waarschijnlijk spelen hierbij ook zoveel andere factoren een rol, dat een uitgesproken oordeel hierover in zijn algemeenheid niet goed mogelijk is. De literatuur geeft onder meer aan dat de motivatie om aan nascholing deel te nemen, behalve door factoren met betrekking tot de inhoud van het programma, ook wordt bepaald door de werkbelasting, de vorm, lokalisatie en de duur van programma's, het zelfvertrouwen van de huisarts, de wens tot contacten met collega's, etc (Cervero, 1981; Pickup et al, 1983; Devitt, 1985; Bloemer & Poiesz, 1987; Branthwaite et al, 1988; Polliack, 1990).

Het invullen van een toets kost ongeveer een kwartier. Deze wetenschap biedt bijvoorbeeld ontwikkelaars van voor de nascholing bestemde onderwijsprogramma's de mogelijkheid hun programma gericht te kunnen plannen.

In het kader van het onderzoek naar de toepasbaarheid van onderwerpgebonden toetsen om op groepsniveau effecten van gerichte nascholing vast te stellen, vond de gegevensverwerking centraal plaats. Overigens geldt voor het gebruik van onderwerpgebonden toetsen als uitsluitend 'smaakmaker' dat centrale verwerking en analyse van de toetsgegevens niet per se noodzakelijk zijn.

DEEL IV (hoofdstuk 10)

VRAAGVORM EN KWALITEIT TOETSEN

De laatste onderzoeksdoelstelling had betrekking op de vraagvorm. In dit gedeelte van het onderzoek stond de vraag centraal wat het effect zou zijn van toepassing van een andere schriftelijke vraagvorm dan de juist-onjuist-vraagtekenvorm in de huisartsgeneeskundige kennistoetsen.

Doelstelling IV:

Bepaling van effecten van de vraagvorm op de huisartsgeneeskundige kennistoetsing.

Vraagstellingen:

IV.1. Welke invloed heeft de vraagvorm op de validiteit van de huisartsgeneeskundige kennistoetsen?

IV.2. Welke invloed heeft de vraagvorm op de betrouwbaarheid van de huisartsgeneeskundige kennistoetsen?

IV.3. Welke invloed heeft de vraagvorm op de praktische toepasbaarheid van de huisartsgeneeskundige kennistoetsen?

Om de vraagstellingen te kunnen beantwoorden is voor één van de onderwerpgebonden toetsen, de kennistoets over 'cholesterol', een variant gemaakt in meerkeuzevorm met drie- en vierkeuze-vragen. Beide toetsvormen werden in random volgorde voorgelegd aan ervaren huisartsen. De resultaten worden in hoofdstuk 10 besproken.

Hoofdstuk 10

INVLOED VAN DE VRAAGVORM OP DE HUISARTSGENEESKUNDIGE KENNISTOETSING

SAMENVATTING

In dit hoofdstuk wordt het resultaat beschreven van een deelstudie die is uitgevoerd naar de inhoud van verschillende vraagvormen die in breed toepasbare huisartsgeneeskundige kennistoetsen kunnen worden gehanteerd. De juist-onjuist-vraagtekenvraagvorm is vergeleken op punten als validiteit, betrouwbaarheid en praktische toepasbaarheid, met traditionele meerkeuzevragen.

Voor de juist-onjuist-vraagteken-kennistoets over 'cholesterol' werd een parallelversie samengesteld met meerkeuzevragen. De toetsen werden voorgelegd aan een aantal huisartsen. De indeling van huisartsen aan een groep die de juist-onjuist-vraagtekenvorm, danwel de meerkeuzevorm voorgelegd kreeg, vond plaats volgens een random procedure. Achtergrondkenmerken van de huisartsen werden verzameld om na te gaan in hoeverre beide groepen vergelijkbaar waren. Daarnaast werden satisfactiegegevens verzameld om na te gaan of de acceptabiliteit van beide vraagvormen van elkaar verschilde.

De resultaten bieden geen duidelijke argumenten voor een keuze voor een van de twee vraagvormen. Het construeren van meerkeuzevragen leverde, dankzij de beschikbaarheid van de vragen in de juist-onjuist-vraagtekenvorm, weinig problemen op, al kon niet altijd worden voldaan aan de wens om vragen met vier alternatieven te formuleren. Voor het toetsen van sommige kenniselementen leek, om inhoudelijke redenen, de meerkeuzevorm geschikter. Voor andere elementen moest echter naar kunstmatige oplossingen worden gezocht om ze in een dergelijke vorm te kunnen formuleren. De betrouwbaarheidsschattingen lieten voor de meerkeuzevorm een iets gunstiger beeld zien dan voor de juist-onjuist-vraagtekenvorm. Een toets bestaande uit meerkeuzevragen vergt echter een langere invultijd dan een toets in de juist-onjuist-vraagtekenvorm. De acceptabiliteit van beide vormen was ongeveer gelijk, al was de non-respons voor de juist-onjuist-vraagtekenvorm groter dan voor de meerkeuzevorm.

10.1. INLEIDING

De literatuur over vraagvormen die geschikt zijn voor gebruik in schriftelijke kennistoetsen met een breed toepassingsgebied, leverde geen pasklaar antwoord voor de keuze voor een bepaalde vraagvorm. In de toetsen die in de voorgaande hoofdstukken zijn beschreven, is aangesloten op de juist-onjuist-vraagtekenvorm die wordt gehanteerd voor de kennistoetsen in de huisartsopleiding. In de huisartsopleiding is om een aantal redenen voor deze vraagvorm gekozen. Met deze vraagvorm kan op een efficiënte wijze worden getoetst, dat wil zeggen dat

in relatief korte tijd over een bepaald onderwerp veel vragen kunnen worden gesteld (Frisbie, 1973). Een praktisch argument was dat met deze keuze werd aangesloten op de toetspraktijk in een aantal huisartsopleidingen en op de voortgangstoetspraktijk in de universitaire artsopleiding in Maastricht, waar de kennistoetspraktijk in de huisartsopleiding van is afgeleid (Kramer & Pollemans, 1990; Kramer et al, 1990; Pollemans & Tan, 1990; Pollemans e.a., 1991).

Van de juist-onjuist-vraagvorm is bekend dat de raadkans groter is dan bij andere schriftelijke meerkeuzevragen (Crocker & Algina, 1986; Dousma & Horsten, 1989; Moelands e.a., 1992). Daardoor zou de betrouwbaarheid in negatieve zin kunnen worden beïnvloed (Suen, 1990). Verandering van de vraagvorm in bijvoorbeeld een traditioneel vaker gebruikte vierkeuzevorm, kan volgens deze theorie leiden tot een hogere betrouwbaarheid en daarmee tot kwalitatief betere toetsen. De betrouwbaarheid is echter niet het enige relevante criterium voor de keuze van een bepaalde vraagvorm. Ook de mogelijkheid met de betreffende vraagvorm de kennis te toetsen die wordt beoogd, de satisfactie van de deelnemers, de noodzakelijke invultijd die een toets vergt in relatie tot het aantal benodigde vragen, en bijvoorbeeld het gemak van de vraagconstructie of gegevensverwerking spelen een rol.

In dit hoofdstuk wordt beschreven wat de invloed is van toepassing van een andere schriftelijke vraagvorm dan de juist-onjuist-vraagtekenvorm op de kwaliteit van de kennistoetsen. De invloed op de validiteit, de betrouwbaarheid en de praktische toepasbaarheid in termen van acceptabiliteit en uitvoerbaarheid van de toetsen komen achtereenvolgens aan de orde.

Onderzoeksvraag IV.1.:

Welke invloed heeft de vraagvorm op de validiteit van de huisartsgeneeskundige kennistoetsen?

Onderzoeksvraag IV.2.:

Welke invloed heeft de vraagvorm op de betrouwbaarheid van de huisartsgeneeskundige kennistoetsen?

Onderzoeksvraag IV.3.:

Welke invloed heeft de vraagvorm op de praktische toepasbaarheid van de huisartsgeneeskundige kennistoetsen?

Opzet onderzoek:

- ontwikkeling parallelkennistoets in meerkeuzevorm voor kennistoets 'cholesterol' in juist-onjuist-vraagtekenvorm
- bepalen verschillen in validiteit, betrouwbaarheid en praktische toepasbaarheid van beide instrumenten

10.2. METHODE VERGELIJKEND ONDERZOEK NAAR VERSCHILLENDE VRAAGVORMEN

10.2.1. Onderzoeksoptzet vergelijkingsstudie

Om kennistoetsing waarin gebruik wordt gemaakt van de juist-onjuist-vraagtekenvorm te kunnen vergelijken met kennistoetsing waarin een andere meerkeuze-vraagvorm wordt toegepast, werd voor de onderwerpgebonden toets over 'cholesterol' een variant ontwikkeld

waarin over dezelfde kenniselementen als in de juist-onjuist-vraagtekenversie, meerkeuzetoetsvragen werden geformuleerd.

Via contactpersonen werd een aantal groepen ervaren huisartsen gevraagd een van beide toetsvormen in te vullen. Toedeling van een toets met de ene of de andere vraagvorm aan de individuele huisartsen gebeurde met behulp van een randomisatietabel. Vervolgens zijn de resultaten die met beide vraagvormen werden behaald, met elkaar vergeleken.

- *opzet validiteitsonderzoek*

Inhoudvaliditeit werd nagestreefd door de toets in de meerkeuzevorm te baseren op dezelfde door de experts op het gebied aangedragen kenniselementen, als de toets in de juist-onjuist-vraagtekenvorm. Om met de vragen ook hetzelfde kennisbegrip te toetsen, werden zoveel mogelijk dezelfde praktijkrelevante huisartsgeneeskundige casusbeschrijvingen als 'stimulus format' gebruikt (van Leeuwen & van Hessen, 1990). De bedoeling was om de vragen in beide toetsen van elkaar te laten verschillen in 'response format' (Swanson & Case, 1992). Als aannemelijk kon worden gemaakt dat elk van beide vraagvormen dezelfde kennisinhoud aan de orde stelde, werd dit beschouwd als ondersteuning voor overeenkomst in de constructvaliditeit van de vragen. De inhoudelijke en contextuele informatie zou in dat geval immers voor beide vraagvormen gelijk zijn (Norman et al, 1987). De vraagconstructeurs werd gevraagd in hoeverre dit uitgangspunt kon worden gehandhaafd bij de feitelijke formulering van de vragen.

Bij de constructie van meerkeuzevragen werd gestreefd naar vragen met vier alternatieve antwoordmogelijkheden. Dit aantal vormt een compromis tussen de wens zoveel mogelijk antwoordmogelijkheden per vraag op te nemen om de raadkans te verkleinen, en de wens de vragen zo efficiënt mogelijk te formuleren (Thorndike, 1982; Haladyna & Downing, 1989-II). In alle alternatieven, inclusief de 'afleiders', de onjuiste alternatieven, werd gevraagd naar elementen die in de casus naar voren kwamen. In de afleiders werd zoveel mogelijk gevraagd naar elementen die volgens een vooronderzoek gerekend mochten worden tot frequent voorkomende misvattingen (van der Weijden, 1992). Bij het formuleren van de vragen werden gangbare constructieregels gehanteerd ten aanzien van de volgorde van de alternatieven en afwisseling van de plaats van het goede antwoord in de reeks alternatieven (Haladyna & Downing, 1989-I; Moelands e.a., 1992).

- *opzet betrouwbaarheidsonderzoek*

In zowel de klassieke testtheorie als de generaliseerbaarheidstheorie wordt ervan uitgegaan dat bepaalde kenmerken van de items die in de toets worden toegepast, mede bepalend zijn voor de te bereiken betrouwbaarheid van de toetsscores. Een van die kenmerken is de raadkans: de kans dat het antwoord goed wordt geraden. Deze kans hangt samen met het aantal antwoordmogelijkheden dat past bij de vraagvorm die wordt gehanteerd (Suen, 1990). Voor elke toets waarin gebruik wordt gemaakt van een gesloten, schriftelijke vraagvorm, geldt dat degenen die de toets beantwoorden, kunnen raden naar het juiste antwoord op de toetsvragen. Te verwachten is dat een deel van de behaalde score door dit raden wordt bepaald. Het raden vergroot het aandeel van de 'error' (de meetfout) in de waargenomen score en verlaagt daardoor de betrouwbaarheid.

De literatuur levert geen pasklare oplossing voor dit probleem. Als wordt aangenomen dat mensen die niet beschikken over de noodzakelijke kennis om een vraag correct te beantwoorden, het antwoord raden, betekent dit voor een meerkeuzevraag met x mogelijke antwoorden, dat de kans op een correct antwoord door raden, gelijk is aan $1:x$. De keuze voor

een antwoord kan echter behalve door een blinde gok, ook bepaald worden door gedeeltelijke kennis, door verborgen prikkels in de tekst van de vraag, of door suggesties, associaties of fouten in de constructie van de vragen. Het raden gebeurt dan niet at random (Crocker & Algina, 1986). Dat betekent dat de werkelijke kans op een correct antwoord hoger kan zijn dan 1:x. Bij meerkeuzevragen met meerdere foute alternatieven (afleiders) speelt ook de aannemelijkheid van de verschillende afleiders een rol. Als men er bij het schrijven van toetsvragen in slaagt dermate aanlokkelijke afleiders te bedenken, dat deze juist attractief zijn voor diegenen met foute of partiële kennis, dan mag men ervan uitgaan dat de kans op goed raden kleiner is dan 1:x (Suen, 1990).

Er bestaan geen afdoende technieken om de invloed van raden op de toetsscores uit te bannen. Daarom wordt er in de praktijk meestal van uitgegaan dat raden at random plaatsvindt, waarbij de kans op goed raden even groot is als de kans op fout raden. Soms wordt gepoogd het effect van raden tegen te gaan door een correctie op de scores toe te passen. De totale score die is behaald wordt dan verlaagd. Deze methode is echter niet onomstreden en eigenlijk alleen gerechtvaardigd bij juist-onjuistvragen (de Gruijter, 1993).

Naarmate het aantal antwoordmogelijkheden per item kleiner is, is de kans groter dat goed wordt geraden. Bij juist-onjuistvragen is de raadkans dus relatief hoog. Daarom wordt er in de literatuur vanuit gegaan dat deze vraagvorm, ten opzichte van andere gesloten schriftelijke vraagvormen, tot minder betrouwbare toetsscores leidt (Moelands e.a., 1992).

De onderwerpgebonden toetsen zowel als de onderwerpen uit de algemene huisartsgeneeskundige kennistoets bleken een grote foutenmarge en daarmee samenhangend een beperkte betrouwbaarheid voor de interpretatie van individuele scores op te leveren (zie hoofdstuk 3 en 8). Om de invloed van de vraagvorm op de betrouwbaarheid te kunnen nagaan, werd daarom de betrouwbaarheid van de cholesteroltoets in de juist-onjuist-vraagtekenvorm vergeleken met die in de meerkeuzevorm. De vraag was hierbij dus met name of verandering van de vraagvorm zou leiden tot een hogere betrouwbaarheid.

De betrouwbaarheid werd geschat conform de op de generaliseerbaarheidstheorie gebaseerde methoden zoals werden uitgevoerd op de algemene huisartsgeneeskundige toets en op de onderwerpgebonden kennistoetsen.

De betrouwbaarheidsberekeningen in dit deel van het onderzoek zijn gebaseerd op de goedscores van de deelnemers. Dat wil zeggen dat geen raadcorrectie is toegepast. Daarvoor waren twee redenen. Correctie voor raden bij meerkeuzevragen wordt in de literatuur niet aanbevolen vanwege de hiervoor vermelde motieven (de Gruijter, 1993). De tweede reden was dat de meerkeuzevorm geen vraagtekenoptie kende. Van het gebruik van de goedscores in de vergelijking tussen beide toetsvormen, mocht worden verwacht dat daarmee zo weinig mogelijk effecten van het gebruik van het vraagteken op de gecorrigeerde score, een rol zouden spelen.

De betrouwbaarheid werd berekend volgens een voor beide toetsvormen identiek *design*, waarin de betrouwbaarheid van individuele toetsscores werd bepaald volgens het zogenaamde 'one facet design' met 'personen', 'items' en de interactie tussen personen en items als variantiebronnen (Cronbach et al, 1972; Brennan, 1983). Voor begripsomschrijvingen en formules wordt weer verwezen naar Bijlage 3. resp. 4.

Variantiebronnen

P (personen)

I (items)

PI (interactie tussen personen en items)

Toelichting:

- personen (P): verschillen tussen personen vormen het object van meting.
- items (I): geeft verschillen in moeilijkheidsgraad aan tussen items.
- personen en items (PI): dit interactie-effect geeft aan in welke mate personen anders worden gerangordend door de verschillende items en is in dit design niet te ontwarren van allerlei andere (fouten)bronnen die de score beïnvloeden.

Voor elke variantiebron werd de relatieve bijdrage aan de totale toetsvariantie berekend en uitgedrukt in een variantiecomponent. De invloed van de verschillende variantiebronnen op de betrouwbaarheid, uitgedrukt in generaliseerbaarheidscoëfficiënten (relatief perspectief) en phi-coëfficiënten (absoluut perspectief), kon zo voor beide toetsvormen zichtbaar worden gemaakt. Ook de bijbehorende standaardmeetfouten (SEM) werden voor beide toetsvormen berekend.

- *opzet onderzoek praktische toepasbaarheid*

De praktische toepasbaarheid werd onderzocht door de ervaringen te inventariseren van de vraagconstructeurs bij het formuleren van meerkeuzevragen. Het construeren van meerkeuzevragen op zich was geen expliciet onderzoeksdoel omdat de toetsvragen als parallelvragen voor de bestaande juist-onjuist-vraagtekenvragen werden geformuleerd.

Daarnaast werd geïnventariseerd hoeveel tijd de huisartsen aangaven nodig te hebben om de meerkeuzevragen dan wel de juist-onjuist-vraagtekenvragen in te vullen. Ook verschillen in satisfactie tussen de groepen die de verschillende vormen kregen voorgelegd, werden geïnventariseerd om zicht te krijgen op eventuele verschillen in acceptabiliteit van beide toetsvormen.

Tenslotte werd de uitvoerbaarheid nagegaan van de verwerkingsprocedures en feedbacksamenstelling voor beide toetsvormen door de benodigde tijd en menskracht te inventariseren.

10.2.2. Onderzoeksinstrumenten

- *kennistoetsen 'cholesterol' met verschillende vraagvormen*

De kennistoets in de juist-onjuist-vraagtekenvorm was dezelfde kennistoets die in hoofdstuk 7 is beschreven. De cholesteroltoets bestond uit 30 juist-onjuist-vraagtekenvragen die betrekking hadden op de fundamentele kenniselementen uit de NHG-Standaard Cholesterol. De items waren alle geschreven in dezelfde casusvorm als was toegepast in de algemene huisartsgeneeskundige kennistoets.

De meerkeuzevorm-toets was een variant hiervan die bestond uit hetzelfde aantal vragen. Over alle kenniselementen die in de vorm van een juist-onjuist-vraagtekenvraag waren gesteld, was het mogelijk een alternatieve vraag in meerkeuzevorm te formuleren (zie Bijlage 17 voor een voorbeeldpagina uit de meerkeuzetoets over cholesterol).

- *vragenlijst achtergrondkenmerken*

Om achteraf te kunnen nagaan in hoeverre de groep huisartsen die de juist-onjuist-vraagtekenvragen had beantwoord afweek van de groep die de meerkeuzevragen had beantwoord, vulden alle deelnemers ook een vragenlijst naar achtergrondkenmerken in. Hiervoor werden dezelfde vragenlijsten gebruikt als in het onderzoek naar de toepasbaarheid van algemene en onderwerpgebonden toetsen in de nascholing (zie Bijlage 6).

- *satisfactievragenlijst*

Ook de satisfactievragenlijst was identiek aan de vragenlijst die werd gebruikt in het onderzoek naar de praktische toepasbaarheid van onderwerpgebonden kennistoetsen (zie Bijlage 13). Zo nodig werd de term 'juist-onjuist-vraagteken' vervangen door 'multiple choice'.

10.2.3. Onderzoeksgroep

De kennistoetsen zijn toegezonden aan 106 ervaren huisartsen. De randomprocedure leidde ertoe dat 56 huisartsen een toets in de meerkeuzevorm kregen, en 50 in de juist-onjuist-vraagtekenvorm.

De respons bedroeg in totaal 61 toetsen (58%): 37 in de meerkeuzevorm en 24 in de juist-onjuist-vraagtekenvorm. De respons op de meerkeuzevorm (66%) was dus hoger dan op de juist-onjuist-vraagtekenvorm (48%). De meeste potentieel deelnemende huisartsen waren niet individueel, maar als (nascholings-)groep geworven via contactpersonen. De toetsen werden door deze contactpersonen in de groepen verspreid en de toetsen werden tijdens groepsbijeenkomsten afgenomen. Er bestond dus een vrij grote kans dat de toetsboekjes ook bij huisartsen die niet aan de toetsing deelnamen, bekend zouden worden. Omdat de toetsen waren gebaseerd op de NHG-Standaard over cholesterol, konden de juiste antwoorden zonder veel moeite worden opgezocht. Daarom is geen poging gedaan om via een herhaalde oproep tot deelname, de respons te vergroten. De al dan niet enthousiasmerende rol van de contactpersonen om de huisartsen te laten deelnemen aan het onderzoek, zal bij de respons ongetwijfeld een rol hebben gespeeld. Praktische overwegingen maakten het echter onmogelijk het effect hiervan in het onderzoek te betrekken.

In tabel 10.1 wordt een overzicht gegeven van de achtergrondkenmerken van de huisartsen die de toets retourneerden.

De groepen komen, behalve naar geslacht, redelijk met elkaar overeen. In de groep die de meerkeuzevragen maakten, bevonden zich naar verhouding wat minder solisten.

De deelnemers zegden toe de toetsen onder examenomstandigheden in te vullen, dat wil zeggen zonder raadpleging van naslagwerken, individueel en onder toezicht. Alle deelnemers ontvingen enige tijd nadat ze de toetsen gemaakt hadden, individueel feedback. Deze was gebaseerd op de geschoonde toets van 29 vragen zoals die ook in het eerder in dit verslag gerapporteerde onderzoek was gehanteerd. De feedback bestond uit een overzicht van de eigen resultaten (voor de juist-onjuist-vraagtekenversie percentage goed, fout, en vraagteken; voor de meerkeuzeversie het percentage goede antwoorden) en van de gemiddelde resultaten van de groep die dezelfde toets had gemaakt (zie Bijlage 18).

Tabel 10.1. Achtergrondkenmerken deelnemers vraagvormonderzoek cholesterol 1993.

Achtergrondkenmerken	JOV-groep huisartsen (N=24)	MCQ-groep huisartsen (N=37)
<i>Geslacht</i> man vrouw	100% -	76% 24%
<i>Leeftijd</i> gemiddeld	44 jaar (std 9)	43 jaar (std 7)
<i>Praktijksituering</i> grote stad stad verstedelijkt platteland platteland onbekend	- 8% 38% 54% -	3% 14% 30% 51% 3%
<i>Werkverband</i> solist duo/associatie groepspraktijk/gezondheidscentrum overig	54% 38% 4% 4%	38% 41% 14% 7%

10.3. RESULTATEN

10.3.1. Validiteit verschillende vraagvormen

Voor alle vragen in de kennistoets met juist-onjuist-vraagtekenvragen konden parallelvormen worden geformuleerd in de meerkeuzevorm. De ervaring met de vraagconstructie leerde dat met de meerkeuzevragen soms iets genuanceerder of precieser naar een bepaald kennisaspect kon worden gevraagd, omdat veel gemaakte fouten als alternatief voor het juiste antwoord konden worden gehanteerd. Bij juist-onjuist-vraagtekenvragen bleven deze aspecten soms impliciet. Dit gold echter zeker niet voor alle vragen. Het was ook niet altijd mogelijk vier gelijkwaardige alternatieven bij een casus op te nemen. In die gevallen werd met drie alternatieven volstaan.

Het was niet goed mogelijk de tekst van de casus die in de juist-onjuist-vraagtekenvorm was gebruikt, integraal in de toets met de meerkeuzevragen over te nemen. Een aantal tekstuele aanpassingen was noodzakelijk om de tekst van de casus en de verschillende antwoorδοpties goed op elkaar te laten aansluiten. Dat betekent dat het 'stimulus format' van beide toetsvormen niet identiek was.

10.3.2. Betrouwbaarheid verschillende vraagvormen

De gemiddelde goedscore op de toets met de juist-onjuist-vraagtekenvorm bedroeg 69%, met een standaarddeviatie van 11%.

Het resultaat van de generaliseerbaarheidsanalyse uitgevoerd op de juist-onjuist-vraagteken-toetsversie wordt weergegeven in tabel 10.2.

Tabel 10.2. *Geschatte variantiecomponenten en percentage van de totale variantie van goedscores op juist-onjuist-vraagteken toets cholesterol (N = 24).*

Variantiebron	Geschatte variantiecomponenten	Percentage van de totale variantie
P	.00530	2.5
I	.03222	15.0
PI	.17755	82.5

De berekeningen die konden worden uitgevoerd op basis van deze gegevens, leverden een relatieve betrouwbaarheid (uitgedrukt in een generaliseerbaarheidscoëfficiënt) voor de goedscores op van .46 met een SEM van 8.

Voor de interpretatie van een individuele goedscore van 69 betekent dit dat de 'ware' goedscore met een zekerheid van 67% ligt tussen 61 en 77, en met een zekerheid van 95% tussen 53 en 85.

De absolute betrouwbaarheid (uitgedrukt in een phi-coëfficiënt) bedroeg .42, met een bijbehorende SEM van 9.

Bij de analyse van de meerkeuzevragen bleek dat op één van de (driekeuze-)vragen ten gevolge van onduidelijkheid in de literatuur (de NHG-Standaard) twee van de drie alternatieven goed moesten worden gerekend. Feitelijk resteerde hier dus een tweekeuzevraag. De gemiddelde goedscore die op de meerkeuzetoets werd behaald, bedroeg 49% met een standaarddeviatie van 13%.

De generaliseerbaarheidsanalyse uitgevoerd over de meerkeuzetoets liet het volgende beeld zien.

Tabel 10.3. *Geschatte variantiecomponenten en percentage van de totale variantie van goedscores op meerkeuzetoets cholesterol (N = 37).*

Variantiebron	Geschatte variantiecomponenten	Percentage van de totale variantie
P	.01111	4.4
I	.04085	16.2
PI	.19973	79.4

Dit leverde een generaliseerbaarheidscoëfficiënt op van .62 met een bijbehorende SEM van 8. Voor een huisarts die de gemiddelde goedscore van 49% behaalde, betekent dit cijfer dat de 'ware' score met een zekerheid van 67% ligt in het gebied tussen 41 en 57, en met een zekerheid van 95% tussen 33 en 65. De absolute betrouwbaarheidsschatting leverde een phi-coëfficiënt op van .57 met een bijbehorende SEM van 9.

In tabel 10.4 worden de consequenties van deze bevindingen voor de interpretatie van toetsscores samengevat. Daarbij wordt uitgegaan van de interpretatie die kan worden gegeven aan de gemiddelde goedscore voor elk van beide toetsvormen. Voor de juist-onjuist-vraagtekenversie bedroeg deze 69%, voor de meerkeuze-versie 49%.

Tabel 10.4. *Verskil in betrouwbaarheidsschattingen van de gemiddelde goedscore voor de juist-onjuist-vraagtekentoets en meerkeuzetoets over 'cholesterol'.*

Vraagvormen	Betrouwbaarheidsschattingen gemiddelde goedscore			
	Relatief perspectief		Absoluut perspectief	
	67%-betrouw- baarheids- interval	95%-betrouw- baarheids- interval	67%-betrouw- baarheids- interval	95%-betrouw- baarheids- interval
Juist-onjuist-vraagteken- vorm (gem. goedscore 69%)	61-77	53-85	60-78	51-87
Meerkeuzevorm (gem. goed- score 49%)	41-57	33-65	40-58	49-67

Tabel 10.4 laat zien dat, ondanks de iets gunstiger betrouwbaarheidscoëfficiënten die de meerkeuze-toetsversie oplevert, de nauwkeurigheid van de gemiddelde scores van beide toetsvormen identiek is. Dat wil zeggen dat voor de interpretatie van de scores van elk van beide toetsvormen met dezelfde onnauwkeurigheid rekening moet worden gehouden.

10.3.3. Praktische toepasbaarheid verschillende vraagvormen

Zoals hiervoor beschreven is de feitelijke vraagconstructie niet direct onderwerp van onderzoek geweest, omdat de meerkeuzevragen werden geconstrueerd rondom dezelfde elementaire kenniselementen als de juist-onjuistvragen. Deze vragen waren bovendien beschikbaar als uitgangsmateriaal voor het formuleren van parallel-meerkeuzevragen. De casuïstiek die als 'stimulus' diende voor de meerkeuzevragen lag daarmee vast. Er is op grond van de ervaring in dit onderzoeksdeel dan ook geen goede uitspraak te doen over de tijd die in acht moet worden genomen voor het construeren van adequate meerkeuzevragen.

Het invullen van een toets van 30 vragen in de juist-onjuist-vraagtekenvorm kostte de deelnemende huisartsen gemiddeld 17 minuten (std 5). Het invullen van de meerkeuzetoets kostte gemiddeld 21 minuten (std 7). Dit verschil was statistisch significant ($p < 0.05$).

In tabel 10.5 wordt het oordeel van de verschillende deelnemersgroepen weergegeven over de bruikbaarheid van beide toetsvormen.

De deelnemers aan de meerkeuzetoets oordeelden gemiddeld wat gunstiger over de bruikbaarheid en de zin van onderwerpgebonden kennistoetsen voor nascholingsdoelen, dan de huisartsen die de juist-onjuist-vraagtekenversie invulden.

De ervaringen met de praktische uitvoerbaarheid van het verwerken van de antwoordformulieren, het analyseren van de gegevens en het samenstellen van feedback voor de meerkeuze-toets, maakten duidelijk dat de benodigde administratieve, methodologische en inhoudelijke deskundigheid daarvoor geheel vergelijkbaar waren met de benodigde handelingen voor de juist-onjuist-vraagtekentoetsen.

Tabel 10.5. *Opvattingen deelnemers (JOV: N = 24; MCQ: N = 37) over toepasbaarheid onderwerpgebonden kennistoetsen (met verschillende vraagvormen) in nascholing.*

Opvattingen	Waardering (in percentages)		
	(geheel) eens	neutraal/geen mening	(geheel) oneens
<i>Vraagvorm is moeilijk</i>			
JOV-deelnemers	36	32	32
MCQ-deelnemers	41	30	30
<i>Toets is nuttig voor opsporen eigen kennislacunes</i>			
JOV-deelnemers	52	24	24
MCQ-deelnemers	70	22	8
<i>Toets is nuttig als 'warming up' bij nascholing</i>			
JOV-deelnemers	76	20	4
MCQ-deelnemers	87	8	5
<i>Toets is nuttig bij nascholingsprogrammakeuze</i>			
JOV-deelnemers	60	24	16
MCQ-deelnemers	70	11	19
<i>Kennistoetsing is niet zinvol</i>			
JOV-deelnemers	12	16	72
MCQ-deelnemers	3	8	89

10.4. CONCLUSIES INVLOED VRAAGVORM

De centrale vraag in dit deel van het onderzoek was of het raadzaam is de vraagvorm in de huisartsgeneeskundige kennistoetsen te veranderen.

Wat betreft de validiteit van de verschillende vraagvormen wordt geconcludeerd dat de inhoudsvaliditeit van beide vormen voldoet. De door experts aangegeven centrale kenniselementen over het thema cholesterol kunnen met beide vraagvormen adequaat worden weergegeven. De inhoud van beide toetsen is dus representatief voor de parate kennis die huisartsen volgens deze experts op het gebied van cholesterol moeten bezitten. Bij de vraagconstructie bleek dat inhoudelijke argumenten ertoe leiden dat soms een voorkeur bestond voor de ene boven de andere vorm. Het was niet altijd mogelijk om voor beide vraagvormen dezelfde 'stimulus' vorm te gebruiken, in termen van inhoudelijke en contextuele informatie in de beschrijving van de huisartsgeneeskundig relevante casuïstiek in de items. Aangenomen moet daarom worden dat beide vraagvormen in constructvaliditeit enigszins van elkaar verschillen.

De gemiddelde goedscores van de huisartsen die hetzij de juist-onjuist-vraagtekentoetsen hetzij de meerkeuzetoetsen retourneerden, verschilden van elkaar. Op de juist-onjuist-vraagteken-toets werden gemiddeld hogere scores behaald dan op de meerkeuzetoets. Deze

bevinding is ook elders gedaan, en hangt waarschijnlijk samen met de per vraagvorm verschillende kans op een goed antwoord bij raden (Frisbie, 1973; Suen, 1990). De gemiddelde scores op de juist-onjuist-vraagtekentoets weken overigens niet af van de gemiddelde scores van de groepen huisartsen die in het kader van het andere onderzoeksdeel dezelfde toets maakten.

Wat betreft de invloed van de vraagvorm op de betrouwbaarheid, lijkt de meerkeuzevorm een wat gunstiger beeld te laten zien dan de juist-onjuist-vraagtekenvorm, al verschillen de gemiddelde goedscores niet in onnauwkeurigheid. Het is onduidelijk of de groepen huisartsen die de verschillende vraagvormen voorgelegd kregen in homogeniteit van elkaar verschilden ten gevolge van de relatief grote non-respons in de groep die de juist-onjuist-vraagtekenvorm voorgelegd had gekregen.

De meerkeuzevragen werden gebaseerd op bestaande juist-onjuist-vraagtekenvragen. Daarom kan, wat betreft de praktische toepasbaarheid van de meerkeuze-vraagvorm, geen goed oordeel worden gegeven over een eventueel verschil in benodigde constructietijd of verschil in gemak wat betreft het construeren van vragen. De deelnemers aan de meerkeuzetoets oordeelden in het algemeen positiever over de bruikbaarheid van kennistoetsen voor huisartsgeneeskundige nascholingsdoelen dan de deelnemers aan de juist-onjuist-vraagtekentoets. Dit ondanks het gegeven dat de toets in de meerkeuzevorm een langere invultijd vergde dan de toets in de juist-onjuist-vraagtekenvorm (gemiddeld ruim een kwartier voor 30 vragen versus gemiddeld ruim 20 minuten). Omdat geen nader onderzoek naar de non-respons is gedaan, is het niet duidelijk in hoeverre deze daadwerkelijk met de vraagvorm samenhang.

De conclusie is dat er geen inhoudelijke argumenten zijn gevonden die variatie met de vraagvorm ontraden. In dezelfde hoeveelheid tijd kunnen echter minder meerkeuzevragen worden beantwoord dan juist-onjuist-vraagtekenvragen. Met dit gegeven zou rekening moeten worden gehouden als de meerkeuzevorm wordt geïntroduceerd bij kennistoetsen die een inhoudelijk breder gebied bestrijken.

Hoofdstuk 11

BESCHOUWING EN DISCUSSIE

SAMENVATTING

Screening, educatie, selectie en evaluatie vormen de belangrijkste gebruiksfuncties die aan kennistoetsen worden toegekend in de huisartsgeneeskundige nascholing. In het onderzoek is nagegaan welke waarde in dit opzicht mag worden toegekend aan een algemene, dat wil zeggen het huisartsgeneeskundige kennisdomein dekkende, toets en aan onderwerpgebonden kennistoetsen.

De algemene huisartsgeneeskundige kennistoets is een goed instrument voor screening van de kennis van groepen huisartsen. De algemene kennis van huisartsen neemt met toenemende leeftijd van de huisartsen af. Het verband dat is gevonden tussen verschillende praktijkkenmerken van huisartsen en hun kennisniveau, wijst erop dat de kennisontwikkeling van huisartsen verband houdt met het soort patiënten, klachten en aandoeningen, die huisartsen in de praktijk tegenkomen. De algemene kennistoets kan door individuele huisartsen goed gebruikt worden als zelfevaluatie-instrument en heeft een functie als leermiddel. De toets is alleen onder voorwaarden bruikbaar als onderdeel van een herregistratie-procedure.

De individuele scores van huisartsen op de onderwerpgebonden toetsen, zijn niet betrouwbaar genoeg. De toetsen, die zijn afgestemd op nascholingsprogramma's, hebben daarvoor op een te beperkt kennisgebied betrekking. Dat neemt niet weg dat de toetsen een goede voorbereiding kunnen vormen op nascholingsprogramma's. Op groepsniveau maken de onderwerpgebonden toetsen duidelijk dat de nascholingsprogramma's wel leiden tot een onmiddellijke kennistoename, maar dat deze toename na verloop van tijd weer afneemt.

Het hoofdstuk wordt afgesloten met beleidsaanbevelingen en aanbevelingen voor nader onderzoek.

11.1. INLEIDING

Dit proefschrift gaat over gebruiksmogelijkheden van kennistoetsen in de huisartsgeneeskundige deskundigheids- en kwaliteitsbevordering. De aanleiding tot het onderzoek waarvan het proefschrift de neerslag vormt, waren een aantal ontwikkelingen, geïnitieerd door de beroepsgroep van huisartsen, die kunnen worden samengevat met de term 'kwaliteitsbevorderingsbeleid'. Het verzamelen van objectieve informatie over het feitelijke competentieniveau, vormt binnen dit beleid een eerste stap in een proces dat moet leiden tot effectieve en efficiënte bewaking en bevordering van de huisartsgeneeskundige zorg. Daarbij gaat het om een vorm van interne kwaliteitsbewaking, die direct is gericht op verbetering van de kwaliteit van de zorg in de dagelijkse praktijk (Buckley, 1989). De beschikbare bevindingen van een

aantal goed uitgevoerde onderzoeken op het gebied van de (huisarts)geneeskundige toetsing en nascholing, geven steun aan de uitgangspunten voor dit beleid:

- competentie is een voorwaarde om 'hogere' doelen zoals gedragsverandering van artsen en verandering in patient outcome-maten, te bereiken (Beaudry, 1989);
- effectief nascholingsonderwijs moet zijn gebaseerd op objectieve bepaling van leerbehoeften van artsen (Davis et al, 1992; Manning & DeBaakey, 1992). Daarnaast spelen ook factoren zoals motivatie en het vermogen verschil te zien tussen het ideale en het feitelijke gedrag, een rol (Geertsma, 1982; Sanazaro, 1983; Anderson et al, 1988).

Dit onderzoek beperkte zich tot de parate kennis die van huisartsen wordt verwacht om hun taken adequaat te kunnen vervullen. Competentie is een begrip dat meer dan alleen kennis omvat (Hays, 1994). Het belang van een voldoende kennisniveau wordt echter ondersteund door recente theorievorming binnen de cognitieve psychologie over de wijze waarop expertise op een bepaald vakgebied ontstaat (Boshuizen, 1989; Schmidt et al, 1990, 1992; Schmidt & Boshuizen, 1993). Volgens deze inzichten vormt de kennisbeheersing op een bepaald terrein, de basis voor het ontwikkelen van probleemoplossende vaardigheden op datzelfde gebied (Norman et al, 1989). Recent onderzoek toont bovendien aan dat scores op schriftelijke toetsinstrumenten een zekere samenhang vertonen met het handelen in de praktijk (Norman, 1991). Met name voor praktizerende artsen geldt daarbij voor de toetsmethode, dat deze wint aan validiteit naarmate de toetsinhoud de werkelijkheid van de dagelijkse praktijk dichter benadert (Miller, 1990; Kane, 1992).

Kennistoetsing in het kader van kwaliteitsbevordering kent de volgende doelen of functies:

- door *screening* van het algemene kennisniveau van huisartsen zouden eventuele verschillen in kennisniveau tussen diverse segmenten van de huisartsenpopulatie zichtbaar gemaakt moeten worden;
- de *educatieve functie* betekent dat toetsgegevens moeten bijdragen aan het inzicht bij huisartsen in hun eigen sterke kanten en kennislacunes;
- de *selectieve functie* betreft de bijdrage van kennistoetsgegevens aan registratie-, c.q. herregistratiebeslissingen;
- de *evaluatieve functie* van kennistoetsen heeft betrekking op de bruikbaarheid van toetsen om effecten van (nascholings-)onderwijs vast te stellen.

Om de toepasbaarheid van kennistoetsen voor deze verschillende functies te kunnen onderzoeken, is onderscheid gemaakt tussen twee vormen van kennistoetsen: *algemene* kennistoetsen, die het gehele huisartsgeneeskundige kennisdomein dekken, en *onderwerpgebonden* kennistoetsen. In dit onderzoek is de kwaliteit van de hiervoor ontwikkelde toetsvormen in termen van validiteit, betrouwbaarheid en praktische toepasbaarheid nagegaan.

De huisartsgeneeskundige kennistoetsen moeten landelijk toepasbaar zijn en zich inhoudelijk richten op de kennis die nodig is om de taken die worden genoemd in het LHV-Basistakenpakket van de huisarts (Springer, 1983), te kunnen uitvoeren. Op onderwerpniveau is een aantal van deze taken vertaald in NHG-Standaarden. Daarop zijn weer deskundigheidsbevorderingspakketten gebaseerd: onderwijspakketten bestemd voor de nascholing. Het is de bedoeling dat uiteindelijk zoveel mogelijk taken worden gedekt in NHG-Standaarden (Rutten & Thomas, 1993). Bij de ontwikkeling van de toetsen is hiermee rekening gehouden. De algemene toets dekt in principe het LHV-Basistakenpakket; de onderwerpgebonden toetsen zijn afgeleid van NHG-Standaarden.

Bij de formulering van de toetsvragen zijn de meest recente inzichten betrokken op het gebied van toetsing en evaluatie op medisch terrein (Newble et al, 1994). Dat betekende onder meer dat de inhoud van de toetsen zo veel mogelijk is afgestemd op praktijkrelevante, 'levensechte' casuïstiek. Veel aandacht is besteed aan de inhoudsvaliditeit van de toetsen, in termen van een zorgvuldige beschrijving van de keuzes en argumenten, die ten grondslag lagen aan de bepaling van de toetsinhouden. Daarnaast is de constructvaliditeit van de kennistoetsen onderzocht. Het kennisbegrip waarvan de toetsvragen een operationalisatie zijn, werd op theoretische gronden aangeduid als *gegeneraliseerde huisartsgeneeskundige ervaringskennis*. Hierin is de conceptuele kennis waarover huisartsen aan het eind van hun opleiding moeten beschikken, vervat. Het is de kennis waarop in beginsel elke huisarts aanspreekbaar is.

11.2. ALGEMENE HUISARTSGENEESKUNDIGE KENNISTOETSING

- screeningsfunctie

De informatie uit het onderzoek naar de validiteit en de betrouwbaarheid van een toets die volgens de in het kader van het onderzoek ontwikkelde blauwdruk is samengesteld, maakt duidelijk welke uitspraken op basis van de toetsscores van ervaren huisartsen kunnen worden gedaan. Uitgaande van een toetslengte van 160 vragen (dat wil zeggen anderhalf tot twee uur toetstijd) geldt dat groepen huisartsen, die bestaan uit meer dan 10 personen, met een hoge geldigheid op grond van hun gemiddelde scores mogen worden gerangordend. Ook uitspraken over de mate van beheersing van de getoetste stof mogen voor dergelijke groepen met een hoge mate van geldigheid worden gedaan. Voor screeningsdoelen voor *groepen* van ervaren huisartsen lijkt de algemene kennistoets dus goed bruikbaar. Dit betekent dat algemene kennistoetsen op landelijk of regionaal beleidsniveau goede diensten kunnen vervullen bij de voorbereiding en planning van de nascholingsprogrammering. Ook kan, door specifieke inhoudelijke vraagstellingen te koppelen aan een dergelijke toets, beleidsondersteunend onderzoek worden verricht. Daarbij valt bijvoorbeeld te denken aan het nagaan hoe het, tien jaar na introductie van NHG-Standaarden, is gesteld met de kennis over die Standaarden bij de Nederlandse huisartsen.

Als men uitsluitend inzicht wil verkrijgen in het gemiddelde kennisniveau van de Nederlandse huisarts, is het op zich niet nodig dat grote groepen huisartsen aan de toetsing deelnemen. Men kan volstaan met een beperkte steekproef van huisartsen bij wie men van tijd tot tijd de kennis toetst. Ook een simpeler design is mogelijk, waarin de toetsonderwerpen random worden verdeeld over een eveneens random gekozen steekproef van huisartsen (Millman & Greene, 1989). Het voordeel van een dergelijk design is dat er minder huisartsen en minder tijd en geld mee zijn gemoeid. Nadeel van dit design is dat het de individuele huisarts weinig oplevert. Deelnemers zouden immers, om selectiebias te voorkomen, verplicht moeten worden tot deelname. Bovendien zou feedback aan de huisartsen in dit design een weinig prominente functie hebben. Als wordt overwogen de algemene screenende kennistoetsing op een dergelijk design te baseren, dan dient opnieuw de praktische toepasbaarheid en met name de acceptabiliteit van de toetsing te worden onderzocht.

Zowel op de totale toets als op onderdelen konden verschillen in gemiddelde kennis worden aangetoond tussen huisartsen in diverse fasen van hun opleiding of ervaring. De hoogste scores werden bereikt aan het eind van de huisartsopleiding. Deze bevinding was conform de

op de literatuurgebaseerde verwachtingen (Neufeld, 1985; Day et al, 1988; Grant & Marsden, 1988; Norcini & Dawson-Saunders, 1994). In de huisartsopleiding wordt, door theoretisch onderwijs en toenemende praktische ervaring, immers het meest geconcentreerd kennis opgedaan. De vragen in de toets zijn geformuleerd door stafleden van de huisartsopleiding. Hoewel ernaar is gestreefd de items zo goed mogelijk te laten aansluiten bij de dagelijkse praktijk, kan dit ertoe hebben geleid dat een aantal vragen voor de ervaren huisarts minder relevant was. Uitbreiding van de groep vragenconstructeurs met huisartsen, die niet intensief zijn betrokken bij de huisartsopleiding, verdient dan ook zeker aanbeveling.

Het verloop van de gemiddelde toetsscores voor klachten- en aandoeninggebonden onderwerpen doet vermoeden dat de kennisontwikkeling daarvan nauw verbonden is met de mate waarin huisartsen in de praktijk met deze onderwerpen worden geconfronteerd. Ook het scorepatroon op de onderdelen van de toets die betrekking hadden op leeftijdscategorieën van patiënten (jongeren, ouderen), spoedeisende gevallen en chronische aandoeningen, wijst erop dat de kennis van huisartsen waarschijnlijk nauw samenhangt met de mate waarin zij met bepaalde problemen in aanraking komen. De resultaten van een onderzoek naar 'soorten van kennis' in de huisartsgeneeskundige kennistoets, wijzen eveneens in deze richting (van Leeuwen et al, in voorbereiding). Voor soorten van kennis waaraan in de opleiding veel aandacht wordt geschonken, zoals kennis over de NHG-Standaarden, worden aan het eind van de huisartsopleiding de hoogste scores aangetroffen. De ervaren huisartsen behalen gemiddeld de hoogste scores op de kennis die al langere tijd onder huisartsen gemeengoed is. Deze bevinding is een onderwerp dat nader onderzoek verdient en vormt wellicht mede een verklaring voor het afnemende algemene kennisniveau van huisartsen bij gemiddeld toenemende ervaring. Mogelijk zakt de kennis waarop niet regelmatig een beroep wordt gedaan, na verloop van tijd weg.

Als het inderdaad zo is dat de kennisontwikkeling van huisartsen verband houdt met het type aandoeningen en het soort patiënten dat zij in hun praktijk al dan niet vaak tegenkomen, moet in de nascholing daarmee rekening worden gehouden. Die zou zich dan vooral moeten richten op aspecten waarmee huisartsen relatief weinig worden geconfronteerd.

- kennistoetsing met educatieve doelen

Een belangrijke functie van toetsing is ondersteuning van het leren door inzicht te bieden in wat men niet weet (Miller, 1976). De feedback op de algemene kennistoets kan goede aanknopingspunten bieden voor het individuele leerproces van huisartsen. De feedback is daartoe onder meer gerubriceerd volgens de op huisartsgeneeskundige morbiditeit gebaseerde hoofdstukken van de ICPC (Schadé & Lamberts, 1989). De toets is daarmee vooral geschikt als instrument voor zelfevaluatie. Door de beschikbaarheid van dergelijke instrumenten kan de nascholing van huisartsen beter worden afgestemd op de principes van volwasseneneducatie (Polliack, 1990; Pietroni, 1992).

Het gebruik van een realistische context voor de presentatie van toetsvragen, zou de toetsing aantrekkelijk maken voor ervaren beroepsbeoefenaren (Norman et al, 1987; Miller, 1990; Swanson et al, 1991). De opinies van de huisartsen die aan de toetsen deelnamen, ondersteunen deze opvatting. De huisartsgeneeskundige relevantie en de zinvolheid van de toetsvragen worden door de deelnemers onderschreven. Vergelijking van de resultaten met de resultaten uit het onderzoek naar de toepassingsmogelijkheden van de algemene kennistoets in de huisartsopleiding, liet zien dat de ervaren huisartsen in het algemeen positiever oordeelden over de algemene kennistoets dan de huisartsen-in-opleiding. Verder werd duidelijk dat, naarmate de deelnemers vaker met een kennistoets worden geconfronteerd, ze

het minder prettig lijken te vinden eraan mee te doen (van Leeuwen, in voorbereiding). Onbekend is of deze bevinding samenhangt met de wijze waarop er binnen de verschillende opleidingen met de toetsen en de resultaten ervan wordt omgegaan, of dat de bevinding samenhangt met het ontstaan van 'toetsmoeheid' of dat andere factoren een rol spelen (Pollemans e.a., 1994).

De gesloten casuïstische vraagvorm lijkt geschikt om praktijkrelevante kennis te meten, op een hoger abstractieniveau dan dat van pure feitenkennis of 'weetjes' (Aiken, 1982). Uit het deelonderzoek naar verschillen tussen de juist-onjuist-vraagtekenvorm en de meer traditionele meerkeuzevorm, kwamen geen ondubbelzinnige argumenten naar voren die pleitten voor een van beide vormen. Inhoudelijke argumenten lijken per vraag de keuze voor de vorm te moeten bepalen. Deze conclusie is in overeenstemming met de literatuur, die aangeeft dat de inhoud van de toetsvragen belangrijker is dan de vorm waarin de vragen worden gegoten (van der Vleuten & Newble, 1994).

Een bijkomend educatief belang van algemene kennistoetsing voor huisartsen, is dat de feedback op de toets over de gemiddelde groepsresultaten, de functie kan hebben van een soort schriftelijke en anonieme 'peergroup'. Deze functie lijkt vooral van belang voor huisartsen in solo-praktijken of huisartsen die anderszins weinig contacten onderhouden met collega's. Regelmatige toetsing van kennis kan bij huisartsen ook leiden tot een nieuwe cultuur, waarin het belang van het bijhouden van het eigen kennisniveau voortdurend wordt onderstreept. Omdat de toetsvragen voorzien zijn van recente en wetenschappelijk verantwoorde documentatie, heeft de toets als bijkomend voordeel dat ze de deelnemers voorziet van een groot aantal gerichte literatuurverwijzingen. In die zin kan de toets ook worden beschouwd als een faciliteit die huisartsen in een tijdsbestek van ongeveer twee uur op de hoogte houdt van de gemiddelde eisen die door de opleiding en de beroepsgroep aan het kennisniveau van huisartsen worden gesteld. Vanuit dit gezichtspunt is de toets zelf een leermiddel. Op basis van deze functie verdient het aanbeveling dat de toets systematisch aan huisartsen beschikbaar wordt gesteld.

Bij de bruikbaarheid van de toets voor individuele huisartsen kunnen kanttekeningen worden geplaatst. De toets is gericht op 'over-all kennis', en geeft geen informatie over specifieke onderwerpen waar iemand bijvoorbeeld heel goed of heel slecht in is. De algemene huisartsgeneeskundige toets is, met andere woorden, geen 'toets-op-maat' die de individuele sterke en zwakke kanten tot in detail bloot legt. Hoewel de toets zo goed mogelijk aansluit op praktisch relevante concepten, is het geen instrument waarin rekening wordt gehouden met typische individuele kenmerken van een bepaalde praktijk (Manning et al, 1986; Manning & DeBakey, 1992). De schriftelijke feedback is persoonlijk en staat niet ter beschikking van derden. De effectiviteit van het instrument op zich in termen van gedragsverandering zal daarom beperkt zijn (Muir Gray, 1986). Deze kanttekeningen gelden overigens voor elke schriftelijke toetsing over een breed vakgebied en vormen geen specifiek bezwaar tegen de algemene huisartsgeneeskundige kennistoets.

Een verwante kanttekening is dat uit recent onderzoek is gebleken dat de diagnostische accuratesse van huisartsen toeneemt met het stijgen van hun ervaring (Hobus, 1994). Deze bevinding lijkt strijdig met de bevinding uit het onderhavige onderzoek dat de gemiddelde kennis van huisartsen afneemt met stijging van de ervaring. Het soort kennis dat in de toets aan de orde wordt gesteld, verklaart waarschijnlijk deze ogenschijnlijke tegenstrijdigheid. Allerlei aspecten die in een schriftelijke kennistoets niet aan bod kunnen komen, zoals visuele of auditieve contextuele informatie, kunnen zeer wel een rol spelen bij het beter leren

diagnostiseren met toenemende ervaring. De toegenomen diagnostische bekwaamheid wijst meer op toename van de toepassing van efficiënte strategieën dan op toename van kennis. Blinde vlekken worden er niet mee onderkend. Dat betekent dat een toenemende diagnostische accuratesse niet strijdig hoeft te zijn met een verminderend algemeen kennisniveau van huisartsen. De implicaties van een verminderend algemeen kennisniveau van huisartsen voor de 'quality of care' verdienen overigens nader te worden onderzocht.

- *kennistoetsing met selectieve doelen*

De betrouwbaarheid van *individuele* scores op de toets van 160 vragen, is te laag (.70) als het erom gaat consequenties te verbinden aan de score voor individuele huisartsen. De marge waarmee rekening moet worden gehouden bedraagt ongeveer 12 scorepunten boven resp. beneden de behaalde score. Hoewel deze foutenmarge geen belemmering vormt voor toepassing van de toets voor educatieve doelen, levert het een probleem op als de toetsresultaten gebruikt worden voor selectiebeslissingen. De relatief lage betrouwbaarheid van de individuele scores betekent dat er een gerede kans bestaat dat bij herhaling van een equivalente toets andere resultaten zouden worden geboekt, zowel wat betreft de onderlinge rangorde van goede en slechte presteerders als wat betreft de absolute hoogte van de scores. Toetsverlenging is een methode die kan leiden tot een hogere betrouwbaarheid. Een langere toets bestaat uit meer vragen, zodat meer inhoudelijke aspecten aan de orde worden gesteld. Daarmee wordt een van de oorzaken van de relatief lage betrouwbaarheid tegengestaan, namelijk het probleem van de inhoudsspecificiteit (van der Vleuten, 1991). Hieronder wordt het empirische gegeven verstaan dat de competentie van een persoon op de ene casus weinig voorspellend is ten aanzien van diens competentie op een andere casus. Dit probleem draagt voor een groot deel bij aan de meetfout en beïnvloedt daardoor de betrouwbaarheid negatief. Het impliceert dat veel casus of vragen nodig zijn om betrouwbare toetsgegevens te verkrijgen. Een hogere betrouwbaarheid, bijvoorbeeld van .80, zoals de literatuur aanbeveelt (Eggen & Sanders, 1993), wordt bereikt bij verlenging van de toets tot minstens 320 vragen (met een bijbehorende toetstijd van 4 uur). Ook in dat geval moet echter rekening gehouden worden met een aanmerkelijke individuele foutenmarge van 8 tot 10 punten plus of min de behaalde score. Het probleem van de inhoudsspecificiteit is niet uniek voor de algemene huisartsgeneeskundige kennistoets, evenmin als het gegeven dat onrealistisch lange toetstijden nodig zijn om het probleem effectief te bestrijden (Norcini & Swanson, 1989).

Een mogelijkheid om het effect van de relatief beperkte betrouwbaarheid van de individuele toetsscores, op de bruikbaarheid van de toets te verminderen, is het combineren van gegevens van opeenvolgende toetsen. Omdat op deze wijze het aantal vragen aanzienlijk wordt uitgebreid, kan de betrouwbaarheid op een aanvaardbaar niveau worden gebracht. Toepassing van deze optie voor ervaren huisartsen is mogelijk als de huisartsen binnen een bepaald tijdsbestek meerdere toetsen maken, bijvoorbeeld binnen een herregistratie-periode van vijf jaar elk jaar een of twee toetsen. Voordeel hiervan is dat ook de ontwikkeling van de kennis in de loop van de tijd kan worden nagegaan. Op die manier kan de huisarts de toets gebruiken als hulpmiddel bij het maken van keuzes in de nascholing. Het totale resultaat kan worden betrokken in herregistratie-beslissingen. In het laatste geval moet er wel rekening mee worden gehouden dat de acceptabiliteit van een dergelijke toets sterk zal veranderen, als deze een meer selecterende rol zou gaan spelen.

Behalve het combineren van scores van opeenvolgende kennistoetsen is het, zeker bij toetsing met selectieve doelen, noodzakelijk kennistoetsen te combineren met toetsen die zich op

andere aspecten van de competentie richten, zoals vaardigheden, kennis over vaardigheden of consultvoering (van der Vleuten et al, 1989; Mol et al, 1992; Hays, 1994; Southgate & Jolly, 1994). Dit laatste voorkomt tevens het risico dat bestaat bij de toetsing van uitsluitend kennis, namelijk dat andere competentie-aspecten in onderwijs of nascholing op de achtergrond raken (Newble & Jaeger, 1983; Frederiksen, 1984; Palincsar & Winn, 1990; Stanley & Al-Shehri, 1993). Verdere ontwikkeling en onderzoek naar de kwaliteiten van andere toetsen dan kennistoetsen is daarom geboden (Tan e.a., 1993; Mol e.a., 1994).

Voor het toekennen van kwalificaties als 'voldoende/onvoldoende' of 'gezakt/geslaagd', met andere woorden een selectief gebruik van de toets, geldt dus een aantal beperkingen. Daarbij moet de plaats van de cesuur (de norm of de grenswaarde) worden betrokken. Uit het onderzoek blijkt dat het wel mogelijk is betrouwbare beslissingen te baseren op een cesuur die ver af ligt van de gemiddelde toetsscore. De keuze voor een cesuur stelt echter tegelijkertijd een validiteitsvraag aan de orde: de beslissing 'voldoende' of 'onvoldoende' op grond van de gekozen cesuur moet ertoe leiden dat inderdaad de huisartsen worden opgespoord die een dermate slechte kennisbeheersing hebben dat het adequaat oplossen van problemen in de praktijk, en daarmee de kwaliteit van de zorg c.q. de gezondheid van de patiënten, erdoor in gevaar komt. Het is echter niet bekend welk minimum-kennisniveau acceptabel is voor het goed kunnen uitoefenen van het huisartsberoep. Daarnaast geldt de eis dat, voor welke cesuur ook wordt gekozen, deze gepaard gaat met een acceptabele foutenmarge voor zak/slaag-beslissingen. Het vaststellen van een acceptabele grens is niet eenvoudig (Mulholland & Harden, 1990).

Nader onderzoek naar de relatie tussen het kennisniveau en de performance of het praktijkgedrag, moet worden uitgevoerd om tot een verantwoorde procedure voor een acceptabele cesuur te komen. De voorspellende waarde van eventueel gecombineerde scores op meerdere toetsen ten aanzien van kwalitatief verantwoord huisartsgeneeskundig handelen, moet in zo'n onderzoek worden betrokken. Pas daarna kan worden vastgesteld of gebruik van de kennistoets in het kader van selectie gerechtvaardigd is. In zo'n onderzoek dienen de consequenties van selectieve toetsing voor ervaren huisartsen te worden betrokken. Ook de gevolgen voor de acceptabiliteit van de toetsing moeten worden onderzocht.

11.3. ONDERWERPGEBONDEN KENNISTOETSING

- kennistoetsing met educatieve doelen

Deelname aan een onderwerpggebonden kennistoets werd door de meeste huisartsen beschouwd als een aangename opstap tot nascholing. De toetsen werden vooral zinvol geacht als individuele voorbereiding op het onderwijsprogramma ('warming up'). Over de rol van specifieke kennistoetsen bij het bepalen van de keuze voor onderwijsprogramma's is men veel minder positief gestemd. Waarschijnlijk spelen hierbij zoveel andere factoren een rol (aanlokkelijkheid van het geboden programma, tijdsduur, plaats etc), dat een oordeel hierover in zijn algemeenheid niet goed mogelijk is. De positieve reacties van de deelnemers op de toetsing in combinatie met de nascholing, vormen een aansporing voor programma-ontwikkelaars om dergelijke toetsen aan hun nascholingsprogramma toe te voegen. Of het onderwijs beter aanslaat door de toetsing voorafgaand aan het programma, moet nader onderzoek aantonen.

- *kennistoetsing met evaluatieve doelen*

Onderwerpgebonden kennistoetsen voor nascholingsdoelen bleken met behulp van huisarts-experts op het betreffende gebied op efficiënte wijze en binnen een beperkte termijn te kunnen worden samengesteld. De NHG-Standaarden en de daarop gebaseerde deskundigheidsbevorderingspakketten vormen een goede basis voor een efficiënte toetsontwikkeling.

De betrouwbaarheid van de onderwerpgebonden toetsen is te laag om uitspraken over het kennisniveau van individuele deelnemers aan de toetsen op te mogen baseren. De precieze betekenis van de feitelijk behaalde score is beperkt. De nascholingsprogramma's bij NHG-Standaarden zijn daarvoor te zeer gericht op een beperkt kennisgebied. Dat betekent dat de toetsen te kort zijn voor betrouwbare meetresultaten voor individuele huisartsen. Deze conclusie heeft consequenties voor de huidige formule die wordt nagestreefd bij de samenstelling van deskundigheidsbevorderingspakketten behorend bij NHG-Standaarden. Deze formule houdt in dat ernaar wordt gestreefd bij elk pakket korte voor- en natoetsen samen te stellen om de groei in kennis op individueel niveau zichtbaar te maken. Dit laatste kan ten gevolge van de te korte lengte van de toetsen, niet worden waargemaakt.

Als er op landelijk of regionaal beleidsniveau, of op individueel niveau, behoefte bestaat aan betrouwbare informatie over de kennis op een specifiek gebied, moet een groter kennisdomein worden aangesproken (zoals bijvoorbeeld 'diabetes mellitus' of 'het bewegingsapparaat'). Uit een goed gevulde centraal beheerde itembank moet dan voor het betreffende onderwerp een toets van voldoende lengte worden samengesteld, die betrouwbare meetinformatie oplevert.

Op groepsniveau blijken de onderwerpgebonden toetsen goed in staat, verschillen aan te tonen in gemiddeld kennisniveau voor en na het volgen van onderwijs. (Kennis)effecten van het onderwijs nemen na verloop van enige maanden af. Nader onderzoek naar inhoud en opzet van onderwijsprogramma's moet meer duidelijkheid geven in de factoren die dit fenomeen beïnvloeden. De bevinding is overigens niet in tegenspraak met de bevindingen die in de literatuur worden gemeld. De uiteindelijke effectiviteit van onderwijsprogramma's moet blijken uit gedragsverandering in de praktijk en verbetering van zorg (Langsley, 1991). Veranderingen daarin vinden plaats door de jaren heen (Wergin et al, 1988). Behalve informatie-overdracht spelen ook andere factoren een rol, met name motivationele (Avorn & Soumerai, 1983; Anderson et al, 1988; Lomas et al, 1989). Het effect van specifieke onderwijsprogramma's lijkt daardoor minder groot (Goldfinger, 1982). Nascholing speelt zich nooit af onder strak gecontroleerde laboratoriumomstandigheden. Alleen al daarom kan niet op basis van één bepaald nascholingsprogramma worden gesteld dat het gedragsverandering bewerkstelligt (Levine & Moore, 1984). Van een eenmalige interventie mag bovendien geen groot effect worden verwacht omdat het in tegenspraak is met de beroepscultuur en werkwijze van de meeste artsen. Deze is eerder behoudend dan gericht op voortdurende verandering (Devitt, 1985; Davis, 1986; Greer, 1988). Nascholing wordt dan ook niet alleen gevolgd om van nieuwe ontwikkelingen op de hoogte te blijven, maar ook om bestaande routines te bevestigen (Lowental, 1981; van der Feen, 1984; Bouhuijs, 1985). Verder hangt het al dan niet veranderen voor een aanzienlijk deel af van allerlei factoren in de werksetting (Grol e.a., 1994).

De literatuur geeft aan dat voor een blijvend effect op de competentie van artsen, bevestiging en bekrachtiging van de informatie nodig is (Goldfinger, 1982; Ferguson et al, 1984; Horder et al, 1986; Kanouse & Jacobi, 1988; Davis et al, 1990). Dit principe werd in het interventie-onderzoek met betrekking tot cholesterol, niet teruggevonden. De groep huisartsen die na de nascholing gedurende enkele maanden een herhaalde persoonlijke praktijkbegeleiding kreeg,

behaalde acht maanden na de nascholing geen hogere gemiddelde score dan de overige in de cholesterolstudie betrokken groepen. Herhaling van de studie met andere onderwerpen moet duidelijk maken in hoeverre deze bevinding themagebonden is.

11.4. AANBEVELINGEN

Op basis van de onderzoeksresultaten worden de volgende aanbevelingen voor beleid en voor nader onderzoek geformuleerd.

11.4.1. Beleidsaanbevelingen

- toetsontwikkeling algemeen

Het verdient aanbeveling de ontwikkeling en implementatie van (kennis)toetsinstrumenten voor gebruik in de deskundigheidsbevordering van huisartsen voort te zetten. Inzet van voldoende huisartsgeneeskundig-inhoudelijk en methodologisch deskundige toetsontwikkelaars is daarvoor noodzakelijk. Zowel voor de ontwikkeling van toetsen, als voor gegevensverwerking en -analyse, en administratieve en technische ondersteuning, is samenwerking door beroepsgroep en huisartsopleiding, zeer gewenst. Daarom wordt aanbevolen deze samenwerking te stimuleren. Uitgaande van de veronderstelling dat de onderlinge toegankelijkheid van vragenbestanden de effectiviteit en de efficiëntie van de toetsontwikkeling bevorderen, verdient de ontwikkeling van een gezamenlijke itembank met kennistoetsvragen bestemd voor opleiding en nascholing daarbij prioriteit.

Verondersteld mag worden dat de inhoudelijke kwaliteit en de acceptabiliteit van kennistoetsen in hoge mate worden bepaald door de kwaliteit van de toetsvragen. Het kunnen beschikken over voldoende, kwalitatief verantwoorde vragen, is daarom van groot belang. Dit wordt het beste gegarandeerd door de productie te spreiden over zoveel mogelijk huisartsen en de controle over de kwaliteit van de vragen aan een vaste groep toetsdeskundigen over te laten.

Informatie over hoe toetsvragen het 'in het veld' gedaan hebben en commentaren van deelnemers aan toetsing zijn van cruciaal belang om de kwaliteit van toetsen te kunnen garanderen. Daarom wordt aanbevolen huisartsgeneeskundige kennistoetsing gepaard te laten gaan met een systematische, continue kwaliteitscontrole van de toetsvragen.

Om de productie van goede toetsvragen te stimuleren, wordt aanbevolen huisartsen hiervoor te belonen, bijvoorbeeld door scholing in het construeren van toetsvragen en de productie van geaccepteerde toetsvragen te honoreren als geaccrediteerde nascholingsactiviteit.

- algemene huisartsgeneeskundige kennistoetsing

Het verdient aanbeveling screenende kennistoetsing bij huisartsen geleidelijk en op basis van vrijwilligheid te introduceren, bijvoorbeeld door systematisch huisartsen gelegenheid tot deelname aan algemene kennistoetsen te bieden.

Het verdient aanbeveling screenende kennistoetsing voor ervaren huisartsen, te introduceren vanuit een beleid dat continue verbetering van de kwaliteit van de zorg nastreeft. Daarin past

het benadrukken van de educatieve functie die de toets en de feedback van de toetsresultaten voor de individuele deelnemers heeft, beter dan het benadrukken van een eventuele selectieve functie van de toets bij herregistratiebeslissingen.

Aanbevolen wordt groepsresultaten van algemene kennistoetsen te gebruiken op landelijk en op regionaal beleidsniveau bij de voorbereiding en planning van nascholingsactiviteiten. Daarbij verdient het ook aanbeveling om gemiddelde groepsscores van huisartsen op onderdelen van de toets te betrekken bij de ontwikkeling van nieuwe NHG-Standaarden.

- *onderwerpgebonden kennistoetsing*

Met betrekking tot de ontwikkeling van onderwerpgebonden kennistoetsen wordt aanbevolen gebruik te maken van de in dit onderzoek gehanteerde methode. Daarbij worden in een consensusprocedure, waarin huisartsen worden betrokken die specifieke deskundigheid hebben op het betreffende onderwerp, de kenniselementen vastgesteld die de inhoudelijke basis voor de toetsontwikkeling moeten zijn.

Aanbevolen wordt regelmatig evaluatie-onderzoek naar de effecten, ook op langere termijn, van nascholing te verrichten met behulp van onderwerpgebonden toetsen.

Aanbevolen wordt onderwerpgebonden toetsen te gebruiken als 'smaakmaker' voor nascholingsprogramma's.

Ten behoeve van de nascholingsprogrammering wordt aanbevolen rekening te houden met kenmerken van de patiëntenpopulatie van de eigen praktijk van huisartsen. Nascholing zou zich met name moeten richten op klachten en aandoeningen die in de eigen praktijk relatief weinig worden gezien.

11.4.2. Aanbevelingen voor verder onderzoek

- *toetsontwikkeling algemeen*

De educatieve functie van het 'vraagteken' bij de schriftelijke, casuïstische kennistoetsvragen in de juist-onjuistvorm is het expliciet maken van zaken die men niet weet. Het gebruik van een 'vraagteken' als antwoord op een vraag waarin een praktisch, huisartsgeneeskundig relevant patiëntenprobleem wordt geschetst, lijkt voor praktizerende huisartsen strijdig met hun dagelijks handelen in de praktijk. Mogelijk is deze functie daarom zinvoller bij toepassing van toetsen in onderwijssituaties dan bij toetsing van praktizerende artsen. Het verdient aanbeveling de zinvolheid van het gebruik van de vraagtekenfunctie te onderzoeken, met name als de toetsen bestemd zijn voor praktizerende huisartsen.

De resultaten uit het onderhavige onderzoek wijzen erop dat toepassing van een andere vraagvorm dan de juist-onjuistvorm in bepaalde gevallen om inhoudelijke redenen de voorkeur verdient. Het verdient aanbeveling om de effecten op de betrouwbaarheid, validiteit en praktische toepasbaarheid te onderzoeken van het gebruik van verschillende vraagvormen voor de algemene huisartsgeneeskundige kennistoets.

Het verdient aanbeveling te onderzoeken in hoeverre de (betrouwbaarheid van) het kennistoetsresultaat verandert, als de toets bij huisartsen onder gecontroleerde omstandigheden (examenomstandigheden) wordt afgenomen onder tijdscontrole.

Het verdient aanbeveling nader onderzoek te verrichten naar de mogelijke consequenties van de uitkomsten van cognitief-psychologisch onderzoek naar de ontwikkeling van expertise, voor de toetsing van kennis van ervaren huisartsen.

In dit kader wordt tevens aanbevolen regelmatig beperkte constructvalideringsonderzoeken te laten uitvoeren naar de geschiktheid van bepaalde vormen van kennistoetsvragen op het gebied van de huisartsgeneeskunde, voor meer en minder ervaren (huis)artsen, en voor bijvoorbeeld internisten.

Aanbevolen wordt om de verdere ontwikkeling van (bestaande) toetsinstrumenten te laten samengaan met een duidelijke onderzoeksvraagstelling naar de kwaliteit van het instrument.

- *algemene huisartsgeneeskundige kennistoetsing*

Aanbevolen wordt de samenhang te onderzoeken van scores die worden behaald op huisartsgeneeskundige kennistoetsen en scores die worden behaald op instrumenten, die zich op andere competentie-aspecten richten, zoals bijvoorbeeld vaardighedentoetsen, of kennis-over-vaardighedentoetsen.

De onderzoeksresultaten wijzen op een mogelijk verband tussen kenmerken van de patiëntenpopulatiekenmerken en de kennis van huisartsen. Aanbevolen wordt deze relatie nader te onderzoeken.

Aanbevolen wordt onderzoek te starten naar de cesaurbepaling voor huisartsgeneeskundige kennistoetsen. Onderdeel van een dergelijk onderzoek zou moeten zijn het verband tussen het kennisniveau en het functioneren in de praktijk. Dergelijk onderzoek moet zich richten op het minimum-kennisniveau dat nog samengaat met acceptabel handelen in de praktijk.

Het gemiddeld afnemend kennisniveau met toenemende ervaring, dat in dit onderzoek is gevonden, verdient nader onderzoek. Mede vanwege bevindingen uit andere onderzoeken die er bijvoorbeeld op wijzen dat de diagnostische accuratesse van huisartsen bij toenemende ervaring wel verbeterd (Hobus, 1994), wordt aanbevolen dergelijk onderzoek mede te richten op de samenhang tussen kennisniveau en kwaliteit van de patiëntenzorg.

- *onderwerpgebonden kennistoetsing*

Met betrekking tot onderwerpgebonden toetsen wordt aanbevolen om deze toe te passen in onderzoek naar de invloed van verschillende onderwijsmethoden op de kennisontwikkeling.

Aanbevolen wordt kennistoetsen te gebruiken voor beleidsondersteunend onderzoek. Als ervan mag worden uitgegaan dat de introductie van NHG-Standaarden minstens tot doel heeft, dat huisartsen kennis nemen van de inhoud van deze Standaarden, zou daarbij kunnen worden gedacht aan het nagaan hoe het, tien jaar na introductie van NHG-Standaarden, is gesteld met de kennis over die Standaarden bij de Nederlandse huisartsen.

SAMENVATTING

Inleiding

De ontwikkeling van toetsingsinstrumenten ten dienste van de nascholing is een kernpunt van het huisartsgeneeskundig kwaliteitsbeleid. Toetsgegevens moeten de individuele huisarts helpen bij de inrichting van diens persoonlijke nascholingstraject. Als artsen vrij zijn hun nascholingsprogramma zonder hulp van dergelijke gegevens vast te stellen, bestaat namelijk het risico dat niet wordt gekozen voor onderwerpen waarvoor nascholing, objectief gezien, het meest nodig is. Zelf ervaren leerbehoeften neigen ertoe aan te sluiten op persoonlijke interesses en specifieke ervaringen. Dit zou ertoe leiden dat vooral wordt gekozen voor nascholingsthema's waar men al veel vanaf weet. Ook op collectief niveau wordt verwacht dat objectieve toetsingsgegevens de nascholing goede diensten kunnen bewijzen. Als immers bekend is wat de 'werkelijke' leerbehoeften en de lacunes van huisartsen zijn, kan dat de planning en de ontwikkeling van nascholingsprogramma's mede sturen. Tenslotte wordt van objectieve toetsingsgegevens verwacht dat ze informatie opleveren die de samenleving voldoende garantie biedt dat de vakbekwaamheid van huisartsen op peil blijft.

Dit onderzoek beperkt zich tot de mogelijkheden die toetsing van huisartsgeneeskundige kennis in deze opzichten te bieden heeft. Een adequaat kennisniveau wordt daarbij beschouwd als één van de voorwaarden om zorg van goede kwaliteit te kunnen leveren.

In het onderzoek stond de vraag centraal naar de waarde en toepasbaarheid van kennistoetsen voor de deskundigheidsbevordering van huisartsen. Om deze vraag te kunnen beantwoorden, moeten de doelen van de toetsing geëxpliciteerd zijn. Van gegevens die worden verzameld met behulp van kennistoetsen op het gebied van de huisartsgeneeskunde, wordt verwacht dat ze het inzicht in sterke en zwakke kanten in het huisartsgeneeskundig kennisniveau vergroten. Daarnaast wordt er een ondersteunende en sturende rol van verwacht bij de planning van (nascholings-)onderwijsactiviteiten. Op basis van deze doelen is een aantal gebruiksfuncties onderscheiden.

- een *screeningsfunctie*: screening van het algemene kennisniveau van huisartsen moet verschillen in kennisniveau tussen diverse segmenten van de huisartsenpopulatie zichtbaar maken, teneinde een gerichtere planning en sturing van de nascholing mogelijk te maken;
- een *educatieve functie* voor de individuele huisarts: dat wil zeggen dat toetsgegevens moeten bijdragen aan het inzicht in sterke kanten en lacunes. Deze functie moet de individuele huisarts helpen de juiste keuzes te maken in het nascholingsaanbod;
- een *selectieve functie* in het kader van de waarde van de bijdrage van kennistoetsgegevens aan registratie-, c.q. herregistratiebeslissingen; en tenslotte
- een *evaluatieve functie* in de zin van de bruikbaarheid van toetsen om effecten van (nascholings-)onderwijs vast te stellen.

Om de toepasbaarheid van kennistoetsen voor deze functies te kunnen onderzoeken, zijn twee soorten kennistoetsen onderscheiden: *algemene* kennistoetsen, die het gehele huisartsgeneeskundige kennisdomein dekken, en *onderwerpgebonden* kennistoetsen.

De algemene kennistoets is erop gericht informatie op te leveren over de kennis over het gehele huisartsgeneeskundige vakgebied teneinde verschillen in kennis tussen individuele huisartsen of groepen van huisartsen te kunnen achterhalen. De algemene toets is daarmee vooral gericht op de screeningsfunctie. Door adequate feedback kan de toets tevens een educatieve functie hebben omdat de toetsresultaten aangeven waar individuele kennislacunes bestaan. Ook voor de selectieve functie geldt dat de toets het kennisdomein moet dekken.

De onderwerpgebonden toetsvorm is erop gericht informatie op te leveren over de kennis over bepaalde huisartsgeneeskundig relevante onderwerpen, c.q. verandering in kennis onder invloed van onderwijs over deze onderwerpen. Ook deze toetsvorm heeft een individuele educatieve functie. Daarnaast is deze toetsvorm op groepsniveau vooral gericht op de evaluatieve functie van kennistoetsing. Voor beide toetsvormen zijn instrumenten ontwikkeld en onderzocht op hun kwaliteit als meetinstrumenten in termen van validiteit, betrouwbaarheid en praktische toepasbaarheid.

In *Hoofdstuk 1* wordt een overzicht gegeven van de achtergronden van het onderzoek en van de literatuur over het gebruik van kennistoetsen in de huisartsgeneeskundige nascholing. Dit mondt uit in de probleemstelling en in de beschrijving van de onderzoeksdoelen. Het hoofdstuk wordt afgesloten met een overzicht van de onderzoeksvragen.

De beschrijving van het onderzoek valt uiteen in vier onderdelen. Deel I beschrijft de algemene huisartsgeneeskundige kennistoetsing, waarvoor is aangesloten op de landelijke kennistoetsontwikkeling binnen de huisartsopleiding. In Deel II wordt beschreven welke samenhangen zijn gevonden tussen kenmerken van huisartsen en hun scores op een algemene kennistoets. Deel III betreft de onderwerpgebonden kennistoetsing. Als exemplarisch voorbeeld zijn kennistoetsen ontwikkeld bij deskundigheidsbevorderingspakketten behorend bij NHG-Standaarden. Daarbij is aansluiting gezocht bij lopende nascholings-activiteiten op het gebied van 'vaginaal bloedverlies' en 'cholesterol'. In Deel IV tenslotte wordt nader ingegaan op de invloed van de vraagvorm op de huisartsgeneeskundige kennistoetsen.

Deel I: Algemene huisartsgeneeskundige kennistoetsing

De *validiteitsstudie* wordt beschreven in *Hoofdstuk 2*. Deze is gericht op de inhouds- en de constructvaliditeit van de toets. De inhoudsvaliditeit van de toets wordt ondersteund door de motivatie en verantwoording van de inhoudsstructuur van de toets. Daartoe is een blauwdruk vastgesteld. Deze moet ertoe leiden dat de toets, vanuit verschillende perspectieven, een evenwichtige afspiegeling vormt van het huisartsgeneeskundig kennisdomein. De blauwdruk is vastgesteld in een stapsgewijze consensusprocedure met experts op het terrein van de huisartsgeneeskundige scholing en toetsing. In de blauwdruk fungeren de huisartsgeneeskundig-inhoudelijke klachten- en aandoeningencomponenten van de International Classification of Primary Care (ICPC) als hoofdingeling. Deze zijn aangevuld met een huisartsgeneeskundig-theoretisch onderdeel. Daarnaast zorgen vier nevenindelingen ervoor dat belangrijke huisartsgeneeskundige invalshoeken, die niet worden gedefinieerd door de ICPC, in voldoende mate in de toets zijn vertegenwoordigd. Dit zijn leeftijdscategorieën van patiënten, aspecten van het consult, spoedeisende gevallen en chronische aandoeningen.

De totale toetslengte is bepaald op 160 vragen, overeenkomend met een geschatte toetsduur van twee uur. De omvang van de blauwdrukcategorieën is vastgesteld op praktische en inhoudelijke gronden. In het onderzoek is gecontroleerd of veranderingen in de relatieve omvang van de hoofdstukken van de blauwdruk van invloed was op de totale toetsbetrouwbaarheid. Dit bleek niet zo te zijn. Dit is beschouwd als een empirische ondersteuning van de inhoudsvaliditeit van de toets.

Om de constructvaliditeit te onderzoeken, is een op basis van de blauwdruk samengestelde kennistoets voorgelegd aan groepen co-assistenten huisartsgeneeskunde, huisartsen-in-opleiding in diverse opleidingsfasen, en aan ervaren huisartsen in verschillende fasen van hun beroeps carrière. Om het kennisbegrip waarover een praktizerend huisarts moet beschikken, goed te kunnen omschrijven, is aansluiting gezocht bij de theorievorming over de ontwikkeling van medische expertise. Deze theorievorming is voortgekomen uit cognitief-psychologische inzichten over de samenhang tussen kennisontwikkeling, het oplossen van problemen en het opdoen van praktijkervaring. Op basis daarvan is het kennisbegrip omschreven als *gegeneraliseerde huisartsgeneeskundige ervaringskennis*. Het is de kennis waarop de huisartsen in beginsel aanspreekbaar zijn. De toetsvragen moeten een beroep doen op huisartsgeneeskundig relevante kenniselementen, ook wel aangeduid als 'key features'. De vragen moeten geloofwaardig zijn en gaan over relevante medische aspecten in de context van patiëntenproblematiek. Als de kennistoets een adequate operationalisatie vormt van dit theoretische kennisbegrip, kon worden verwacht dat het kennisniveau, uitgedrukt in gemiddelde toetsscore, van groepen artsen in verschillende fasen van opleiding en ervaring, geleidelijk hoger zou worden. Ervan uitgaande dat de ontwikkeling van gerichte kennis over huisartsgeneeskundig-inhoudelijke concepten het meest geconcentreerd plaatsvindt tijdens de huisartsopleiding, werd verwacht dat het kennisniveau het hoogst zou zijn aan het eind van de huisartsopleiding. Ervaren huisartsen bleken als groep inderdaad gemiddeld significant hogere scores te behalen dan de groep huisartsen-in-opleiding, die op hun beurt weer significant hogere gemiddelde scores behaalden dan de co-assistenten huisartsgeneeskunde. De ervaren huisartsen bereikten een gemiddelde toetsscore die ongeveer op het niveau ligt dat door huisartsen-in-opleiding vlak voor het einde van de huisartsopleiding wordt bereikt. Deze bevindingen worden beschouwd als ondersteuning van de constructvaliditeit van de algemene huisartsgeneeskundige kennistoets.

De betrouwbaarheid maakt duidelijk hoeveel 'ruis' of meetfouten in de score zijn verdisconteerd. In *Hoofdstuk 3* wordt ingegaan op de *betrouwbaarheid* van de scores. De betrouwbaarheid is vanuit drie perspectieven geïnterpreteerd: relatief, absoluut en beslissingsgeoriënteerd. Het relatieve perspectief is vooral van belang bij gebruik van de toets voor screening en educatie. Het absolute perspectief is van belang als het gaat om absolute beheersing van de getoetste stof, dat wil zeggen als het gaat om de vraag of aan de scores een absolute betekenis kan worden toegekend. Het beslissingsperspectief speelt een rol bij de betrouwbaarheid van zak-slaag beslissingen, dat wil zeggen bij selectie of (her)registratie. Omdat de generaliseerbaarheidstheorie het mogelijk maakt verschillende betrouwbaarheidsindices te berekenen voor alle onderscheiden perspectieven, is het onderzoek gebaseerd op deze theorie. Alle berekeningen zijn gebaseerd op goed-min-foutscores, uitgedrukt in percentages van de maximaal te bereiken score.

De gemiddelde toetsscore van de huisartsen is 45% met een standaarddeviatie van 11%. De betrouwbaarheid van individuele toetsscores vanuit relatief perspectief bedraagt .70 met een standaardmeetfout van 6. Dat betekent een foutenmarge (95%-betrouwbaarheidsinterval) voor

de individuele toetsscores van 12 punten die moeten worden opgeteld of afgetrokken van de individuele score. Vanuit absoluut perspectief bedraagt deze marge plus of min 14 punten. Verdubbeling van de toets tot 320 vragen (vier uur toetstijd) levert een betrouwbaarheid op van .80 voor beide perspectieven, maar de marge (95%-betrouwbaarheidsinterval) bedraagt dan nog steeds plus of min 8, respectievelijk 10. Betrouwbare zak-slaag beslissingen (met een betrouwbaarheidscoëfficiënt van .80) kunnen worden genomen op basis van een cesuur die ligt onder een goed-min-foutscore van 35% of boven die van 55%. De betrouwbaarheid van gemiddelde scores liet een veel gunstiger beeld zien. Het 95%-betrouwbaarheidsinterval bedroeg 2, resp. 6 punten voor het relatieve en het absolute perspectief, bij een groepsgrootte van ongeveer 30 huisartsen.

Op basis van de betrouwbaarheidsschattingen wordt geconcludeerd dat de kennistoets voor screeningsdoelen voldoende betrouwbare gegevens oplevert. Ook voor educatieve doelen is de toets goed bruikbaar. Bij de interpretatie van individuele scores moet rekening worden gehouden met een te grote foutenmarge om zak-slaagconsequenties aan het toetsresultaat te kunnen verbinden. De betrouwbaarheid van beslissingen varieert met de plaats van de cesuur en is hoger naarmate deze verder van de gemiddelde score af ligt. Omdat onduidelijk is welk minimumniveau aan kennis nog verenigbaar is met het goed functioneren als huisarts, is op grond van dit onderzoek geen goede uitspraak te doen over de beste plaats van de cesuur voldoende/onvoldoende.

Op de *praktische toepasbaarheid* van algemene huisartsgeneeskundige kennistoetsing wordt ingegaan in *Hoofdstuk 4*. Een op de blauwdruk gebaseerde toets is voorgelegd aan ruim 350 ervaren huisartsen en hen is gevraagd naar hun opinie over de kwaliteit en de toepasbaarheid van de toets. De huisartsen oordeelden in het algemeen positief over de toets en achtten de toets huisartsgeneeskundig relevant. De huisartsen vonden de toets het beste bruikbaar als instrument voor zelfbeoordeling. Het invullen van een toets van 160 vragen kost gemiddeld ongeveer anderhalf uur, hetgeen vanwege het feit dat bijna eenderde van de deelnemende huisartsen de toets te lang vond, het maximaal haalbare lijkt.

Wat betreft de toetsconstructie en organisatie van grootschalige toetsafnames, wordt geconcludeerd dat de benodigde constructietijd (gemiddeld 70 minuten per item, uitgevoerd door een ervaren huisarts-toetsvragenschrijver) en de voorwaarden die een de gegevensverwerking en feedbackverzorging vergen, vereisen dat deze worden uitgevoerd door een centrale organisatie met een adequate outillage.

Deel II: Kennis in relatie tot achtergrondkenmerken van huisartsen

Het verband tussen de algemene kennis van huisartsen en hun achtergrondkenmerken wordt weergegeven in *Hoofdstuk 5*. Dit verband is onderzocht door de gemiddelde toetsscores van ervaren huisartsen te relateren aan een aantal persoonskenmerken, demografische en professionele kenmerken.

Naarmate huisartsen meer ervaring hebben, is hun gemiddelde kennistoetsscore lager. Alleen lidmaatschap van het NHG en opleiderservaring vertonen een significant positief verband met de gemiddelde kennistoetsscore. Er kon geen relatie worden vastgesteld tussen werkverband of praktijksituering en kennisniveau.

De scores voor met name medisch-inhoudelijke thema's doen vermoeden dat de kennis daarvan nauw samenhangt met de mate waarin huisartsen er in de praktijk mee worden geconfronteerd. Dit geldt bijvoorbeeld voor aandoeningen van het bewegingsapparaat of

huidaandoeningen. Hetzelfde wordt gevonden voor 'medicamenteus beleid'. Het vermoeden dat kennis over bepaalde onderwerpen en daadwerkelijke praktijkervaring met deze thema's samenhangen, wordt ondersteund door de bevinding dat huisartsen met praktijken met relatief veel ouderen, gemiddeld significant hogere scores behalen op de casuïstiek, die betrekking heeft op 'ouderen (boven 75 jaar)' en 'chronische aandoeningen'. Huisartsen met veel jonge gezinnen in de praktijk scoren gemiddeld wel hoger dan hun collega's op de categorie 'jongeren beneden 15 jaar', maar dit verschil is niet significant.

Hoofdstuk 6 betreft het verband tussen de zelfkennis van huisartsen en hun scores op de algemene kennistoets. De toetsscores van de huisartsen zijn daartoe vergeleken met het oordeel dat ze over hun eigen expertise gaven op de onderwerpen uit de kennistoets. Gemiddeld gaven de huisartsen ruim twee expertisegebieden aan. Samenhang tussen het aantal expertisegebieden en de gemiddelde toetsscore kon niet worden vastgesteld. Ruim 40% van de huisartsen (en zelfs ruim 70% van de vrouwelijke huisartsen) gaf aan zich meer deskundig te achten dan collega's op het gebied van psychosociale aandoeningen. Deze groep huisartsen behaalde op dit terrein echter geen significant hogere scores dan de overige huisartsen. Ook de vrouwelijke huisartsen behaalden geen hogere scores dan hun mannelijke collega's op dit gebied. De conclusie is dat er geen eenduidige relatie is gevonden tussen kennisscore en inschatting van eigen expertise: de totale gemiddelde scores van de huisartsen hangen niet significant samen met het aantal gebieden waarop de huisartsen zich deskundig achten. Als wordt gekeken naar specifieke onderwerpen, scoren de huisartsen die zich expert noemen op die onderwerpen meestal op die gebieden wél iets hoger dan hun collega's die zich geen expert voelen. De verschillen zijn echter meestal niet significant en het verband ligt ook wel eens andersom.

Deel III: Onderwerpgebonden huisartsgeneeskundige kennistoetsen

De *validiteit* van onderwerpgebonden kennistoetsen die zijn bedoeld voor gebruik in de nascholing van huisartsen, wordt beschreven in *Hoofdstuk 7*. De validiteitsstudie was gericht op het aantonen van de inhoudsvaliditeit en de constructvaliditeit van deze toetsen. Inhoudsvaliditeit is nagestreefd door de inhoud van de toetsen zo goed mogelijk af te stemmen op de kenniselementen die door deskundigen op het betreffende gebied wordt beoordeeld als de parate kennis waarover de praktizerend huisarts moet beschikken over het betreffende onderwerp. Voor de kennistoets over cholesterol werd een speciale procedure toegepast waarbij de fundamentele kenniselementen werden vastgesteld in een consensusprocedure, waarbij de samenstellers van de NHG-Standaard Cholesterol werden betrokken. De constructvaliditeit van de toetsen is onderzocht door *paralleltoetsen* voor te leggen aan huisartsen, direct voorafgaand aan en meteen na afloop van nascholing over het betreffende thema. Voor het onderwerp 'cholesterol' is daarnaast een retentiemeting verricht ongeveer acht maanden na het nascholingsprogramma. In de constructvalidering van de toets over cholesterol was ook een vergelijkingsgroep betrokken. Bovendien deden niet alleen huisartsen maar ook huisartsen-in-opleiding mee aan de toetsing en nascholing over cholesterol. De gemiddelde voorttoetsscore van huisartsen is, zoals verwacht, significant hoger dan die van de huisartsen-in-opleiding. Direct na de nascholing bereiken beide groepen een even hoge gemiddelde score. De gemiddelde score op de retentietoets van de groepen die de nascholing hadden gevolgd, is significant hoger dan hun aanvankelijke score, al is de score in vergelijking met de score onmiddellijk na de nascholing wel gedaald. De gemiddelde score

van de vergelijkingsgroepen verandert niet. Uit een stapsgewijze multiële regressie-analyse blijkt dat alleen de score op de voortoets en het al dan niet bijgewoond hebben van het nascholingsprogramma, een significante invloed hebben op de bereikte retentiescore. De conclusie uit de validiteitsstudie is dat het goed mogelijk is inhoudsvalide onderwerpgebonden kennistoetsen te construeren waarmee op groepsniveau verschillen kunnen worden gemeten die samenhangen met deelname aan een nascholingsprogramma. Deze laatste bevinding ondersteunt de constructvaliditeit van de toetsen.

Hoofdstuk 8 handelt over de *betrouwbaarheid* van de onderwerpgebonden toetsen. Deze is onderzocht op basis van dezelfde theoretische uitgangspunten als de algemene kennistoets. Conform de verwachting blijkt de betrouwbaarheid zowel vanuit het relatieve als het absolute perspectief te laag om consequenties aan individuele scores te kunnen verbinden. Op individueel niveau bieden de onderwerpgebonden kennistoetsen geen betrouwbare informatie over de kennisbeheersing. Dit staat overigens een zinvol educatief gebruik van de toetsen niet in de weg. De gemiddelde scores zijn betrouwbaar genoeg om uitspraken te kunnen doen over de evaluatieve functie van de onderwerpgebonden toetsen. Met de toetsen is het op groepsniveau goed mogelijk om kennisverschillen in samenhang met het volgen van nascholingsprogramma's vast te stellen.

De *praktische toepasbaarheid* van onderwerpgebonden kennistoetsing wordt beschreven in *Hoofdstuk 9*. Deze is onderzocht door het oordeel van de doelgroep, praktizerende huisartsen, in te winnen over de waarde van het gebruik van kennistoetsen bij nascholingsprogramma's. Daarnaast is evenals bij de algemene huisartsgeneeskundige kennistoetsing, de uitvoerbaarheid van de toetsing nagegaan door de tijd en menskracht te inventariseren die nodig is om toetsen te construeren, door de huisartsen te vragen hoeveel tijd het invullen van een toets kost en door hun oordeel over de zin en bruikbaarheid van dergelijke toetsen in te winnen, en tenslotte door na te gaan in hoeverre analyse en verwerking van de toetsgegevens en feedbacksamenstelling praktisch uitvoerbaar zijn. De beschikbaarheid van de NHG-Standaarden, vooral als daarnaast kan worden beschikt over een lijst met essentiële kenniselementen uit de Standaarden, blijkt een efficiënt hulpmiddel bij de vraagconstructie. De huisartsen beoordelen kennistoetsing bij nascholing vrijwel unaniem als een zinvolle activiteit, vooral als smaakmaker of 'warming up' voor de nascholing. Onderwerpgebonden kennistoetsen zijn volgens de gebruikers niet zo geschikt als hulpmiddel bij de keuze van bepaalde nascholingsprogramma's. Wat betreft de verwerking en analyse van de toetsgegevens en het samenstellen van feedback geldt, evenals is geconcludeerd bij de algemene kennistoets, dat dit onder de voorwaarde dat adequate gegevensverwerkingse apparatuur en programmatuur beschikbaar is, geen problemen oplevert.

Deel IV: Vraagvorm in huisartsgeneeskundige kennistoetsen

Het laatste onderzoeksdeel betreft een studie die is uitgevoerd naar de invloed van verschillende vraagvormen die in breed toepasbare huisartsgeneeskundige kennistoetsen kunnen worden gehanteerd, op de kwaliteit van het meetinstrument. Het resultaat wordt beschreven in *Hoofdstuk 10*. De invloed is nagegaan van het gebruik van de juist-onjuist-vraagtekenvraagvorm op de validiteit, de betrouwbaarheid en de praktische toepasbaarheid van kennistoetsen, in vergelijking met het gebruik van traditionele meerkeuzevragen. Daartoe is voor de juist-onjuist-vraagteken-kennistoets over cholesterol een parallelversie samengesteld

met meerkeuzevragen. Beide toetsen zijn voorgelegd aan ervaren huisartsen. De indeling van huisartsen aan de groep die hetzij de juist-onjuist-vraagtekenvorm, hetzij de meerkeuzevorm voorgelegd kreeg, vond plaats volgens een random procedure. Achtergrondkenmerken van de huisartsen zijn verzameld om na te gaan in hoeverre beide groepen vergelijkbaar waren. Daarnaast is nagegaan of de acceptabiliteit van beide vraagvormen verschillend is.

De resultaten wijzen er niet op dat een keuze voor een van beide vormen evidente voordelen biedt. Voor het toetsen van bepaalde kenniselementen lijkt om inhoudelijke redenen de meerkeuzevorm geschikter. Voor andere elementen moet echter naar kunstmatige oplossingen worden gezocht om ze in de vorm van een vierkeuzevraag te kunnen formuleren. De betrouwbaarheidsschattingen laten voor de meerkeuzevorm een iets gunstiger beeld zien dan voor de juist-onjuist-vraagtekenvorm. Een toets bestaande uit meerkeuzevragen vergt echter een langere invultijd dan een toets in de juist-onjuist-vraagtekenvorm. De acceptabiliteit van beide vormen is ongeveer gelijk, al was de non-respons voor de juist-onjuist-vraagtekenvorm groter dan voor de meerkeuzevorm. Geconcludeerd wordt dat er geen argumenten zijn om, als dat op inhoudelijke gronden wenselijk is, niet te variëren met de vraagvorm. Deze conclusie is in overeenstemming met de literatuur die aangeeft dat met betrekking tot de inhoudelijke en psychometrische kwaliteit van schriftelijke kennistoetsen de vorm ondergeschikt is aan de inhoud.

Conclusie

In *Hoofdstuk 11* worden de onderzoeksresultaten aan een nadere beschouwing onderworpen. Daarbij wordt vooral ingegaan op de bruikbaarheid van de kennistoetsgegevens voor de verschillende gebruiksfuncties die in de inleiding zijn geformuleerd. Geconcludeerd wordt dat de algemene huisartsgeneeskundige kennistoets een goed instrument lijkt voor screening op groepsniveau. Wat betreft de educatieve functie is de toets als zelfevaluatie-instrument en als leermiddel zeer bruikbaar voor huisartsen. De betrouwbaarheid van de individuele toetsscores is bij de huidige toetslengte te laag om de toets zonder meer toe te passen in het kader van selectie- of herregistratiebeslissingen. Wat betreft de onderwerpgebonden toetsen wordt geconcludeerd dat de kennisdomeinen, die de NHG-Standaarden en bijbehorende deskundigheidsbevorderingspakketten bestrijken, te beperkt zijn om er kennistoetsen op te baseren die voldoende betrouwbare resultaten opleveren om er op individueel niveau conclusies aan te kunnen verbinden. De toetsen kunnen echter een nuttige rol vervullen in de voorbereiding op nascholingsprogramma's. De onderwerpgebonden toetsen laten zien dat effecten van nascholing in termen van kennisverandering na verloop van tijd lijken weg te ebben. Dit gegeven leidt tot de aanbeveling dergelijk onderzoek voor verschillende onderwerpen te herhalen. Daarnaast moet worden nagegaan in hoeverre verschillende onderwijsvormen verschillende effecten bewerkstelligen. Onderwerpgebonden kennistoetsen vormen een goed instrument om dergelijk onderzoek mee uit te voeren.

SUMMARY

Introduction

Quality assurance in general practice has many facets. One of these is the development of assessment instruments aimed at enhancing the efficiency and efficacy of continuing education in general practice. One of the goals of assessment procedures is to help the individual general practitioner in making the right choices in continuing medical education. If doctors chose topics for continuing medical education of their own accord, there is a considerable risk of their choosing topics that are unnecessary from an objective point of view. Self-perceived learning needs tend to match personal interests and specific experiences. Another application for objective tests lies in the field of planning and developing of continuing medical education. The use of collective test results can enhance the quality of this process. If the 'true' learning needs and gaps in knowledge of general practitioners are known, planning and development of educational programmes can be better guided. Finally, objective test results can be used to show society as a whole that the professional ability of general practitioners is being maintained at a high level.

We have studied the various goals of assessment for one area of the competence of general practitioners, i.e. that of knowledge. An adequate level of knowledge is considered an important prerequisite for quality of care in general practice.

Our main research question concerns the suitability and applicability of knowledge tests in continuing medical education for general practitioners. To answer this question testing purposes must be clear. Objectively collected information should increase the insight to strengths and weaknesses in general practice knowledge. Besides, it should support and guide the planning of educational activities in continuing education. Based on these purposes, we distinguished the following areas of application of knowledge tests:

- *screening*: obtaining information about the general knowledge level in various segments of the population of general practitioners to enhance planning and guiding of continuing education;
- *education*: giving individual general practitioners feedback on strengths and weaknesses to help them make the right choices from the educational programmes offered;
- *selection*: knowledge tests as an element in certification and recertification decisions; and
- *evaluation*: assessing the usefulness of knowledge tests in establishing the effectiveness of continuing education programmes.

To meet these goals two types of knowledge tests have been developed: a *general* test covering the whole field of general practice, and *specific* knowledge tests covering certain specific topics. The general test is meant both for screening of the whole population and for giving feedback to individual general practitioners. To serve selective purposes the test should also cover the whole domain of general practice.

The specific tests can be used for individual feedback about one's knowledge of specific topics. It can also be used at group level to assess the effectiveness of a particular continuing medical education course. The validity, reliability and applicability of both the general and the specific test have been assessed.

Chapter 1 gives an overview of the background of the study and of the literature on knowledge testing in continuing education for general practice, culminating in the problem definition and research questions.

The study is made of four parts. Part I is a description of the construction and validation of the general knowledge test. This part of the study has been carried out in close cooperation with a study on the usefulness and applicability of the general knowledge test in vocational training for general practice. In Part II the relation between characteristics of general practitioners and their scores on the general knowledge test is described. The specific tests are presented in Part III. Two subjects were chosen in order to study the validity, reliability and applicability of these tests, 'vaginal bloodloss' and 'cholesterol'. These knowledge tests are related to national standards of the Dutch College of General Practitioners. In Part IV the effects of varying the item format in knowledge tests for general practice is reported.

Part I: General knowledge test

Chapter 2 covers the content and construct *validity* of the test. To guarantee high content validity a blueprint was developed for the construction of the test. This blueprint ensures a balanced representation of the domain of general practice knowledge from different perspectives. The blueprint was developed in a stepwise consensus procedure by experts in the field of continuing education and assessment in general practice. The chapters of the International Classification of Primary Care (ICPC) concerning symptoms and complaints, and diagnoses and diseases, were adopted as the main classification in the blueprint. A chapter on theoretical topics of relevance to general practice, covering non-clinical aspects, was added. Furthermore, four additional classifications ensure the representation of important issues in general practice which are not defined in the ICPC. These are 'age of patients', 'aspects of consultation' (diagnostics, management), 'emergency medicine' and 'chronic diseases'. The total test consists of 160 items, resulting in an average testing time of two hours. The size of the blueprint chapters was set using criteria of practicality as well as content. Reliability studies, in which the relative sizes of all chapters were varied, showed no influence of chapter size on reliability, which supports the high content validity of the test in yet another way.

A general knowledge test, constructed according to the blueprint, was taken by medical students in their clerkships, trainees during vocational training for general practice at various levels, and experienced general practitioners in the various phases of their professional careers. Theories from the field of cognitive psychology on changes in knowledge and problem solving skills on gaining experience in daily practice, were used to specify the knowledge concept in the test. We have described it as *generalized expert knowledge on general practice*. This is the knowledge a general practitioner is assumed to master at the end of vocational training. The items in the knowledge test should focus on 'key features', knowledge elements that are crucial to general practice care. Furthermore, the items should cover relevant aspects of realistic patient problems.

If indeed the knowledge test is an adequate representation of this theoretical knowledge concept, it was expected that the mean test scores would gradually increase with training and/or experience. As the acquisition of specific knowledge of relevant concepts of general practice takes place in a more intensive manner during general practice training than thereafter, we postulated that the scores would be highest at the end of training. Indeed, students during their clerkships had the lowest scores and the mean scores of all experienced general practitioners surpassed the mean scores of all trainees in general practice. The general practitioners had an average score at the level of the group of trainees just before the end of vocational training. These findings support the construct validity of the general knowledge test.

The reliability of a test elucidates the amount of measurement error which has been taken into account in the scores. *Chapter 3* covers the *reliability* of the test scores of the general knowledge test from three perspectives: relative (norm), absolute (domain) and decision (mastery) oriented. For screening and educational purposes the relative-oriented perspective is the most relevant. When one is interested in how well the scores actually represent mastery of the knowledge, the absolute-oriented perspective is paramount. Lastly, for pass-fail decisions, such as those for (re)certification, the decision-oriented perspective is used. Because generalizability theory provides a framework to estimate reliability indices according to all three perspectives, the reliability study is based on this theory. All estimates are based on correct-minus-incorrect scores, expressed as percentages of the maximum score.

The average test score for general practitioners is 45% with a standard deviation of 11%. The reliability of individual test scores from a relative perspective is .70 with a standard error of measurement (SEM) of 6. This means that the 95%-confidence interval of an individual score lies between 12 points added to or subtracted from the individual score. From the absolute perspective this interval is plus or minus 14 points. Doubling the test size to 320 items (four hours testing time) increases the estimated reliability coefficient to .80 from both perspectives but still leaves one with a 95%-confidence interval of plus or minus 8 to 10 points. Reliable pass-fail decisions (with a generalizability coefficient of .80) can be taken with either a fail score below 35% or a pass score over 55% (correct-minus-incorrect).

The reliability of group mean scores is far better, with a 95%-confidence interval of 2 and 6 points for the relative and absolute perspective respectively, assuming a group size of about 30 general practitioners.

In conclusion, the reliability of test scores for screening purposes is sufficient. This is also the case when the scores are used for educational purposes. However, the reliability of the individual scores is not high enough for pass-fail decisions. The reliability of pass-fail decisions varies according to where the cut-off score is placed and is higher, the further the cut-off score is from the average score. As there is no evidence so far as to which level of knowledge is a prerequisite for the good performance of a general practitioner in daily practice, it is impossible to pronounce upon the best place for a cut-off score for pass-fail decisions.

Chapter 4 describes the *applicability* of the general knowledge test. The knowledge test based on the blueprint was taken by more than 350 general practitioners. They were asked for their opinion of the quality and the applicability of the test. On the whole, they were positive about the test, and considered it relevant to daily practice. They thought the test most useful as an instrument for self-evaluation. One-third of the participants considered the time it took to take

the test (1½ to 2 hours) too long. This limits the feasibility of increasing the number of items per test. As to the organization of test construction and test taking, the following conclusions can be drawn. With an average time of 70 minutes for the construction of one test item, when done by an experienced general practitioner-test writer, and considering the ample resources needed for data processing and compiling and returning of feedback, the use of this test is only feasible on a large scale when done by a central organization with extensive resources and equipment.

Part II: Knowledge related to background characteristics of general practitioners

The relationship between the knowledge of general practitioners and their background characteristics is described in *Chapter 5*. Personal, demographic and professional characteristics of general practitioners were related to their knowledge as measured with the general test. As general practitioners gain more experience their average score gradually decreases. Membership of the Dutch College of General Practitioners and experience as a trainer were the only two variables that had a significant positive correlation with the test score. No relation was found for type of practice or geographical situation and test score. Especially the scores on patient-bound themes suggest that this kind of knowledge is closely related to the degree with which general practitioners are confronted with the themes in their daily practice, for example musculoskeletal or skin problems. The same was found for the part of the test that concerned 'medication'. The assumption of a relation between knowledge on certain issues and actual practice experience with these issues, is supported by the finding that general practitioners with relatively many elderly patients, achieve significantly higher mean scores on items on 'elderly patients (age over 75)'. The mean score of general practitioners with relatively more young families in their practice is higher than the mean score of their colleagues on items on 'young people (age under 15)', but the difference is not statistically significant.

Chapter 6 describes the relationship between self-knowledge of general practitioners and their scores on the general knowledge test. The test scores of the general practitioners are compared with their judgment on their own expertise on topics of the test. On average the general practitioners reported specific expertise on more than two subjects. No relation was found for the number of 'expertise subjects' and mean test score. More than 40% of the general practitioners (even more than 70% of the female general practitioners) reported being more expert than their colleagues on the subject of psychosocial problems. This group of general practitioners, however, did not achieve significantly higher scores on this subject than the other general practitioners. Moreover, the female general practitioners did not achieve higher scores on this subject than their male colleagues.

The conclusion is that no unambiguous relationship was found between test scores and the assessment of own expertise. The average score of 'experts' is mostly slightly higher than the score of 'non-experts' on a certain subject. However, the differences were seldom significant and opposite relationships were found as well.

Part III: Specific general practice knowledge tests

The *validity* of specific knowledge tests meant for use in the continuing education of general practitioners is described in *Chapter 7*. The validity study was aimed at establishing content

validity and construct validity. Content validity has been sought by attuning the content of the tests as closely as possible to the knowledge elements, which experts in the area judged as the necessary knowledge relevant to the day-to-day practice of a general practitioner. A special procedure was adopted for the development of the knowledge test on cholesterol. The essential knowledge elements were determined in a consensus procedure which included the composers of the standard on cholesterol. The construct validity of the tests has been investigated by submitting parallel tests to general practitioners, immediately before and directly after continuing education on the theme of the test. On top of this a follow-up post test eight months after the continuing education, was done for the theme 'cholesterol'. The design of the study on construct validity of the cholesterol test involved the use of comparison groups. In addition, not only general practitioners but also trainees in general practice were involved in the education and testing on cholesterol. The mean test score of the general practitioners before the education programme took place was, as expected, significantly higher than the mean score of the trainees. Immediately after following the educational programme both groups achieved the same mean test score. The mean follow-up test score of the groups that followed the educational programme was significantly higher than their mean score at the onset. The mean score after eight months showed a decrease in comparison to the score after following the programme. The mean scores of the comparison groups did not change. A stepwise multiple regression analysis showed that only the initial scores and the participation in the educational programme had a significant influence on the score after eight months. The conclusion from the validity study is that the construction of a content-valid specific knowledge test on a particular topic is feasible. Also group differences related to the participation in continuing education programmes can be measured with this test. This last finding supports the construct validity of the tests.

Chapter 8 concerns the *reliability* of the specific tests. The investigation of the reliability of the tests was based on the same theoretical assumptions used in the overall knowledge test. As expected, reliability turns out to be too low to attach consequences to individual test scores, from the relative as well as from the absolute perspective. On an individual level the specific knowledge tests do not give reliable information. This does, however, not hinder meaningful educational use of the tests. The mean group scores are reliable enough to warrant the use of the specific tests in evaluation studies. The tests make it readily possible to bring about changes in knowledge related to continuing education on group level.

The *applicability* of specific knowledge tests is described in *Chapter 9*. This aspect was investigated by obtaining the judgment of the practising general practitioners on the practical value of knowledge tests in continuing education. Furthermore, the feasibility of testing was verified by surveying time and manpower necessary for the construction of tests, by asking the general practitioners how much time was needed to fill in a test, and finally, by investigating the feasibility of data processing and the production of feedback.

The availability of the national standards, and the list with essential knowledge elements distilled from the standards proved to be an efficient aid to the construction of test items. Nearly all general practitioners found the testing of knowledge in combination with continuing education worthwhile, especially as a 'warming-up' to the education programme. They did not find the specific knowledge tests very useful in helping to choose between continuing educational programmes. As to data processing and production of feedback, the same

conclusions as those of the overall knowledge test apply. Provided adequate data processing hardware and software are available, there are no important impediments.

Part IV: Question format of general practice knowledge test

The final part of our investigation concerns a study of the influence of two different question formats that can both be used in broadly applicable general practice knowledge tests, on the quality of the instruments. The results are described in *Chapter 10*. The influence of the application of the true-false-question mark item format, compared to the application of traditional multiple choice questions, on validity, reliability and practical applicability of knowledge tests is examined. A parallel version to the true-false-question mark version of the cholesterol test was constructed in the multiple choice format. Both test versions were taken by experienced general practitioners. General practitioners were randomly allocated to groups that got the true-false-question mark version or the multiple choice version. Differences in the acceptability of both question formats were investigated.

The results do not indicate that a choice for one of both formats yields obvious advantages. For content related reasons the multiple choice format seems more appropriate for the testing of certain knowledge elements. However, for some other elements the multiple choice format requires artificial alternative choices. Reliability estimates show a slightly better picture for the multiple choice format compared to the true-false-question mark format. However, it takes longer to do a test that consists of multiple choice items than a test in the true-false-question mark format. The acceptability of both formats is similar, but the non-response for the true-false-question mark format was higher than for the multiple choice format. The conclusion is that no proof has been found that variation in question format is unwise, if content-related arguments make variation advisable. This conclusion is in line with the literature which indicates that regarding the psychometric quality of tests, the format of written knowledge tests is of secondary importance to the content.

Conclusion

In *Chapter 11* the overall results of the research are examined more closely. Attention focuses on the usefulness of the knowledge test results for the different purposes that have been distinguished. The conclusion is that the overall general practice knowledge test seems an adequate instrument for the screening of groups. The test is very useful for general practitioners as an instrument for self-evaluation and as an educational tool. The reliability of individual test scores using the current test is too low to apply the test as part of selection or recertification decisions. Concerning the specific tests, the conclusion is that the knowledge areas as covered by the national standards and their corresponding educational programmes are too limited to function as a basis for knowledge tests large enough to be sufficiently reliable to allow conclusions on an individual level. However, the tests may be very useful in the preparation of continuing education. The specific tests indicate that the effects of continuing education in terms of knowledge changes seem to fade away with the lapse of time. This fact leads to the recommendation to repeat this type of evaluation research for other themes. Besides that, differences in impact of different educational methods should be investigated. Specific knowledge tests form an adequate instrument to carry out this type of research.

LITERATUUR

- Aiken LR. Writing multiple-choice items to measure higher-order educational objectives. *Educ Psychol Measmt* 1982; 42: 803-6.
- Alexander PA, Judy JE. The interaction of domain-specific and strategic knowledge in academic performance. *Rev Educ Res* 1988; 58: 375-404.
- Allal L. Generalizability theory. In: Keeves JP, editor. *Educational Research, Methodology and Measurement. An International Handbook*. Oxford: Pergamon Press, 1988.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association, 1985.
- Anderson CM, Chambers S, Clamp M. et al. Can audit improve patient care? Effects of studying use of digoxin in general practice. *Br Med J* 1988; 297: 113-4.
- Anderson J. *The Multiple Choice Question in Medicine*. Tunbridge Wells: Pitman Medical Publishing Co Ltd, 1976.
- Anderson J. For multiple choice questions. *Med Teach* 1979; 1: 37-42.
- Anonymus. The essential nature of continuing medical education [editorial]. *Med Educ* 1987; 21: 85.
- Arocha JF, Patel VL. Question-answering in relation to assessment of problem-solving. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, editors. *Teaching and Assessing Clinical Competence*. Groningen: BoekWerk Publ, 1990.
- Asbaugh D, McKean R. Continuing Medical Education: the philosophy and use of audit. *JAMA* 1976; 13: 1485-8.
- Astin AW, Panos RJ. The evaluation of educational programs. In: Thorndike RL, editor. *Educational Measurement (second edition)*. Washington DC: American Council on Education, 1971.
- Avorn J, Soumerai SB. Improving drug therapy decisions through educational outreach: a randomized controlled trial of academically-based 'detailing'. *N Engl J Med* 1983; 308: 1457-63.
- Axt-Adam P, Wouden JC van der, Hoek H, Does E van der. Het effect van nascholing op het aanvragen van laboratoriumdiagnostiek door huisartsen. *Huisarts Wet* 1993; 36: 451-4.
- Beadry JS. The effectiveness of continuing medical education: a quantitative synthesis. *J Contin Educ Health Prof* 1989; 9: 285-307.
- Bender W. Statistiek en meten: wat moet je daarover weten? Item-analyse. *Bulletin Medisch Onderwijs* 1994; 13: 37-43.
- Benson JA. Certification and recertification: one approach to professional accountability. *Ann Intern Med* 1991; 114: 238-42.
- Bereiter C. Some persisting dilemmas in the measurement of change. In: Harris CW, editor. *Problems in Measuring Change*. Madison: University of Wisconsin Press, 1963.
- Berg AO. Does continuing medical education improve the quality of medical care? A look at the evidence. *J Fam Pract* 1979; 8: 1171-4.

- Berkel HJM van. De diagnose van toetsvragen [Dissertatie]. Amsterdam: Centrum voor Onderzoek van het Wetenschappelijk Onderwijs (COWO) Universiteit van Amsterdam, 1984.
- Berkel HJM van, Bax AE, redactie. Beoordelen in het onderwijs. Een handleiding voor het construeren van toetsen en het evalueren van leerdoelen en onderwijsvormen. Almere: Versluijs, 1990.
- Bertram DA, Brooks-Bertram PA. The evaluation of Continuing Medical Education: a literature review. *Health Educ Monogr* 1977; 5: 330-62.
- Berwick DM. Continuous improvement as an ideal in health care. *N Engl J Med* 1989; 320: 53-6.
- Besseling CWM, Tromp JHM, Vleuten CPM van der, Verbraeck A. De ontwikkeling van voortgangstoetsen in het hoger gezondheidszorgonderwijs. Eindrapport Samenwerkingsverband. 's Gravenhage: Haagse Academie voor Fysiotherapie, Deventer Academie voor Fysiotherapie, Rijksuniversiteit Limburg, 1987.
- Binsbergen JJ van, Brouwer A, Drenth BB van, Haverkort AFM, Prins A, Weijden T van der. NHG-Standaard Cholesterol. Huisarts Wet 1991; 34: 551-7.
- Bloemer J, Poiesz ThBC. Nascholingsbehoeften en -wensen bij de Nederlandse huisarts. Verslag van een consumentenonderzoek. Tilburg: Katholieke Universiteit Brabant, Sectie Economische Psychologie, 1987.
- Boeringa FH, Sluijs EM. Bibliography no. 45. Recertification of Primary Health Care Professionals. Utrecht: NIVEL, 1993.
- Bordage G, Grant J, Marsden P. Quantitative assessment of diagnostic ability. *Med Educ* 1990; 24: 413-25.
- Bordage G, Page G. An alternative approach tot PMP's: the 'key features' concept. In: Hart IR, Harden RM, editors. *Further Developments in Assessing Clinical Competence*. Montreal: Heal Publications, 1987.
- Bordage G, Zacks R. The structure of medical knowledge in the memories of medical students and general practitioners: categories and prototypes. *Med Educ* 1984; 18: 406-16.
- Boshuizen HPA. De ontwikkeling van medische expertise. Een cognitief-psychologische benadering [Dissertatie]. Maastricht: Rijksuniversiteit Limburg, 1989.
- Boshuizen HPA, Schmidt HG. Kennisontwikkeling en onderwijs in de geneeskunde: een subtiel evenwicht. *Bulletin Medisch Onderwijs* 1993; 12: 98-109.
- Boshuizen HPA, Schmidt HG, Coughlin LD. On-line representation of a clinical case and the development of expertise. Paper presented at the AERA meeting. Washington, 1987.
- Boshuizen HPA, Schmidt HG, Coughlin LD. On the application of medical basic-science knowledge in clinical reasoning: implications for structural knowledge differences between experts and novices. In: *Proceedings of the 10th Annual Conference of the Cognitive Science Society*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1988.
- Boud D. The role of self-assessment in student grading. *Assessment and Evaluation in Higher Education* 1989; 14: 20-30.
- Bouhuijs PAJ. Planning continuing medical education for general practitioners: a case study from the Netherlands. *Studies in Higher Education* 1985; 10: 269-75.
- Branthwaite A, Ross A, Henshaw A, Davie C. Continuing education for general practitioners. Occasional paper 38. London: Royal College of General Practitioners, 1988.
- Brennan RL. Elements of Generalizability Theory. Iowa City, Iowa: American College Testing Publications, 1983.

- Brennan RL, Kane MT. An index of dependability for mastery tests. *J Educ Meas* 1977; 14: 277-88.
- Brown S. Trends in assessment. In: Harden RM, Hart IR, Mulholland H, editors. *Approaches to the Assessment of Clinical Competence*. Dundee: Centre for Medical Education, 1992.
- Buckley EG. Quality assessment or quality control? *J R Coll Gen Pract* 1989; 39: 309-12.
- Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally Publ. Company, 1966.
- Carmine EG, Zeller RA. *Reliability and validity assessment*. Beverly Hills, London: Sage Publications, 1985.
- Casparie AF. Kwaliteit in de gezondheidszorg. Huidige inzichten en toekomstige ontwikkelingen. *Medisch Contact* 1989; 44: 477-82.
- Cauffman JG, Rasgon IM, Mayne JC, Schubert WF, Forsyth RA, Lenoski EF. Relationship between quality of CME instruction and changes in physicians' patient-management plans. *J Med Educ* 1985; 60: 486-8.
- Cervero RM. A factor analytic study of physicians' reasons for participating in continuing education. *J Med Educ* 1981; 56: 29-34.
- Chi MTH, Glaser R. The measurement of expertise: analysis of the development of knowledge and skill as a basis for assessing achievement. In: Baker EL, Quellmalz ES, editors. *Educational Testing and Evaluation. Design, analysis and policy*. Beverly Hills, London: Sage Publications, 1980.
- Chi MTH, Glaser R, Farr MJ, editors. *The Nature of Expertise*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1988.
- Claessen HFA, Boshuizen HPA. Recall of medical information by students and doctors. *Med Educ* 1985; 19: 61-7.
- Cohen G, Davis D. Continuing Medical Education and the competent family physician. *Can Fam Physician* 1986; 32: 348-51.
- Colliver JA, Verhulst SJ, Williams RG, Norcini JJ. Reliability of performance on standardized patient cases: a comparison of consistency measures based on generalizability theory. *Teach Learn Med* 1989; 1: 31-7.
- Commissie Nascholing NHG. *Nascholing maken. Criteria voor nascholing van huisartsen (2e druk)*. Lelystad: Meditekst, 1989.
- Corbett TC. The impact of continuing medical education on quality of care: implications of the CME literature. In: LeBreton PP, Bryant VE, Zweigig DL, et al., editors. *The evaluation of continuing education for professionals: a systems view*. Seattle: University of Washington, 1979.
- Crocker L, Algina J. *Introduction to classical and modern test theory*. Fort Worth, etc. Orlando, Florida: Holt, Rinehart and Winston, Inc, 1986.
- Cronbach LJ. Five perspectives on the validity argument. In: Wainer H, Braun HI, editors. *Test Validity*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1988.
- Cronbach LJ. Construct validation after thirty years. In: Linn RE, editor. *Intelligence: Measurement, Theory and Public Policy*. Urbana: University of Illinois Press, 1989.
- Cronbach LJ, Furby L. How we should measure 'change' - or should we? *Psychol Bull* 1970; 74: 68-80.
- Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley, 1972.
- Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull* 1955; 52: 281-302.

- Davis D. Evaluating continuing medical education: common sense and science. *Can Med Assoc J* 1986; 134: 485-6.
- Davis D, Haynes RB, Chambers L, Neufeld VR, McKibbin A, Tugwell P. The impact of CME: a methodological review of the Continuing Medical Education literature. *Eval Hlth Prof* 1984; 7: 251-83.
- Davis DA, Norman GR, Painvin A, Lindsay E, Ragbeer MS, Rath D. Attempting to ensure physician competence. *JAMA* 1990; 263: 2041-2.
- Davis DA, Thomson MA, Oxman AD, Haynes RB. Evidence for the effectiveness of CME. A review of 50 randomized controlled trials. *JAMA* 1992; 268: 1111-7.
- Day SC, Norcini JJ, Webster GD, Viner ED, Chirico AM. The effect of changes in medical knowledge on examination performance at the time of recertification. In: *Proceedings of the 27th Annual Conference on Research in Medical Education*. Chicago, Illinois: Association of American Medical Colleges, 1988.
- Devitt JE. Common sense in continuing medical education. *Can Med Assoc J* 1985; 133: 1119-20.
- Dixon J. Evaluation criteria in studies of continuing education in the health professions: a critical review and a suggested strategy. *Eval Hlth Prof* 1978; 1: 47-65.
- Donabedian A. The definition of quality and approaches to its assessment. Ann Arbor, Michigan: Health Administration Press, 1980.
- Dousma T, Horsten A. Tentamineren. Groningen: Wolters/Noordhoff, 1989.
- Drenth PJD. Inleiding in de testtheorie. Deventer: Van Loghum Slaterus, 1975.
- Drenth PJD, Sijsma K. Testtheorie. Inleiding in de theorie van de psychologische test en zijn toepassingen. Houten/Zaventem: Bohn, Stafleu, van Loghum, 1990.
- Dubois V, Everwijn S, Geldorp G van, et al. The construction of a new curriculum of vocational training for general practice in The Netherlands. Utrecht: College voor Huisartsgeneeskunde, 1987.
- Düsman H. Landelijke experimentele kennistoets [intern rapport]. Utrecht: Samenwerkingsverband-IOH, 1991.
- Düsman H. Over de betrouwbaarheid van kennistoetsen geconstrueerd volgens de blauwdruk (2e versie) [intern rapport]. Utrecht: Samenwerkingsverband-IOH, 1992.
- Düsman H. Betrouwbaarheid onder andere toetsomstandigheden [intern rapport]. Utrecht: Samenwerkingsverband-IOH, 1992.
- Düsman H. Betrouwbaarheid van specifieke toetsen voor praktiserende huisartsen. Verwachtingen gebaseerd op beschikbare gegevens [intern rapport]. Utrecht: Samenwerkingsverband-IOH, 1992.
- Düsman H. Rapportage inter-indelaarsbetrouwbaarheidsonderzoek. Utrecht: Samenwerkingsverband-IOH, 1993.
- Ebel RL. Must all tests be valid? *Am Psychol* 1961; 16: 640-7.
- Ebel RL. Proposed solutions to two problems of test construction. *J Educ Meas* 1982; 19: 267-78.
- Ebel RL. The practical validation of tests of ability. *Educ Meas, issues pract* 1983; 2: 7-10.
- Eekhof JAH, Mol SSL, Pielage JC. Hoe vaak is soms: een onderzoek naar de betekenis van semi-kwantitatieve termen in toetsvragen. In: *Vleuten CPM van der, Scherpbier AJJA, Pollemans MC, redactie. Gezond Onderwijs-I*. Houten/Zaventem: Bohn, Stafleu, van Loghum, 1992.

- Eekhof JAH, Pollemans MC, Verwijnen GM. Toetsbeoordelingscommissie. In: Metz JCM, Scherpbier AJJA, Vleuten CPM van der, redactie. Medisch Onderwijs in de Praktijk. Een gids voor docenten. Assen: van Gorcum & Comp, (in voorbereiding).
- Eggen TJHM, Sanders PF. Psychometrie in de praktijk. Arnhem: Cito Instituut voor Toetsontwikkeling, 1993.
- Escovitz, GH, Davis D. A bi-national perspective on continuing medical education. Acad Med 1990; 65: 545-50.
- Evered DC, Williams HD. Postgraduate education and the doctor. Br Med J 1980; 280: 626-8.
- Everwijn SEM, Bomers GBJ, Knubben JA. Een competentiegericht managementcurriculum. De overbrugging van de kloof tussen kennisverwerving en het toepassen van kennis. Tijdschrift voor Hoger Onderwijs 1990; 8: 55-66.
- Everwijn SEM, Palm J. Bruikbare kennis en vaardigheden: een visie op onderwijs. Opleiding en Ontwikkeling 1993; 11: 13-9.
- Ewell PT. To capture the ineffable: new forms of assessment in higher education. Review of Research in Education 1991; 17: 75-125.
- Eijk JThM van, Gubbels JW. Wetenschappelijk onderzoek in de huisartsgeneeskunde. NHG-publikaties Nr. 1 (2e herziene druk). Lelystad: Meditekst, 1987.
- Feen JAE van der. Evaluatie van nascholing. Warffumcursussen voor huisartsen. Medisch Contact 1984; 39: 413-6.
- Feldt LS, Brennan RL. Reliability. In: Linn RL, editor. Educational Measurement (third edition). Washington DC: National Council on Measurement in Education/American Council on Education, 1989.
- Feltovich PJ, Barrows HS. Issues of generality in medical problem solving. In: Schmidt HG, Volder ML de, editors. Tutorials in problem-based learning; a new direction in teaching the health professions. Assen: van Gorcum, 1984: 128-42.
- Ferguson KJ, Caplan RM, Williamson PS. Factors associated with behavior change in family physicians after CME presentation. J Med Educ 1984; 59: 662-6.
- Fleming PR. The profitability of 'guessing' in multiple choice question papers. Med Educ 1988; 22: 509-13.
- Forrest JM, McKenna M, Stanley IM, Boaden NT, Woodcock GT. Continuing education: a survey among general practitioners. Fam Pract 1989; 6: 98-107.
- Foulkes J. Combining components of assessment. In: Newble D, Jolly B, Wakeford R, editors. The Certification and Recertification of Doctors. Issues in the Assessment of Clinical Competence. Cambridge: Cambridge University Press, 1994.
- Frederiksen N. The real test bias. Influences of testing on teaching and learning. Am Psychol 1984; 39: 193-202.
- Frisbie DA. Multiple choice versus true-false: a comparison of reliabilities and concurrent validities. J Educ Meas 1973; 10: 297-304.
- Frisbie DA. The effect of item format on reliability and validity: a study of multiple choice and true-false achievement tests. Educ Psychol Measmt 1974; 34: 885-92.
- Geertsma RH, Parker RC, Whitbourne SK. How physicians view the process of change in their practice behavior. J Med Educ 1982; 57: 752-61.
- Gellhorn A. Periodic physician recredentialing. JAMA 1991; 265: 752-5.
- Gick ML, Holyoak KJ. The cognitive basis of knowledge transfer. In: Cormier SM, Hagman JD, editors. Transfer of learning: contemporary research and applications. San Diego: Academic Press, 1987.
- Glaser R. Education and thinking: the role of knowledge. Am Psychol 1984; 39: 93-104.

- Glaser R. Towards new models for assessment. *Intern J of Educ Research* 1990; 14: 475-83.
- Goldfinger SE. Continuing medical education: the case for contamination. *New Engl J Med* 1982; 306: 540-1.
- Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991; 66: 762-9.
- Gordon MJ. Self-assessment programs and their implications for health professions training. *Acad Med* 1992; 67: 672-9.
- Grant J, Marsden P. The structure of memorized knowledge in students and clinicians: an explanation for diagnostic expertise. *Med Educ* 1987; 21: 92-8.
- Grant J, Marsden P. Primary knowledge, medical education and consultant expertise. *Med Educ* 1988; 22: 173-9.
- Green K. Multiple choice and true-false: reliability and validity compared. *J Exp Educ* 1978; (..): 42-4.
- Greer AL. The state of the art versus the state of the science. The diffusion of new medical technologies into medical practice. *Int J Technol Assess Health Care* 1988; 4: 5-26.
- Groen GJ, Patel VL. Medical problem-solving: some questionable assumptions. *Med Educ* 1985; 19: 95-100.
- Groen GJ, Patel VL. The relationship between comprehension and reasoning in medical expertise. In: Chi MTH, Glaser R, Farr MJ, editors. *The Nature of Expertise*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1988.
- Grol RPTM. Naar een 'kwaliteitssysteem' in de huisartsgeneeskunde [oratie]. Utrecht: Nederlands Huisartsen Genootschap, 1991.
- Grol R. Kwaliteitssystemen in de huisartsgeneeskunde: wat betekent dit voor de huisarts? *Huisarts Wet* 1993; 36: 106-12.
- Grol RPTM, Everdingen JJE van, Casparie AF. Invoering van richtlijnen en veranderingen. Utrecht: De Tijdstroom, 1994.
- Grol R, Everwijn S, Dubois V. De meerjarige beroepsopleiding tot huisarts. IV: Opzet van het curriculum. *Medisch Contact* 1987; 42: 278-82.
- Grol R, Wensing M, Jacobs A, Baker R, editors. *Quality assurance in general practice. The state of the art in Europe*. Utrecht: Nederlands Huisartsen Genootschap, 1993.
- Grol R, Zwaard A. Deskundigheidsbevordering van huisartsen. NHG-publicaties Nr 3. Utrecht: Nederlands Huisartsen Genootschap, 1990.
- Grol R, Zwaard A. Deskundigheids en -kwaliteitsbevordering van huisartsen. Het opzetten en uitvoeren van programma's ter verbetering van de zorg. Utrecht: Nederlands Huisartsen Genootschap, 1990.
- Grosswald SJ. Designing effective educational activities for groups. In: Green JS, Grosswald SJ, Suter E, Walthall DB, editors. *Continuing Education for the Health Professions*. San Francisco, etc.: Jossey-Bass Publ, 1984.
- Gruijter DNM de. Correctie voor gokken bij tentamens. *Onderzoek van Onderwijs* 1993; 22: 54-5.
- Haan M de, Lisdonk EH van de, Voorn ThB. De kern van de huisartsgeneeskunde. Utrecht: Bunge, 1992.
- Haladyna TM, Downing SM. A taxonomy of multiple-choice item-writing rules. *Appl Meas Educ* 1989; 2: 37-50.
- Haladyna TM, Downing SM. Validity of a taxonomy of multiple-choice item-writing rules. *Appl Meas Educ* 1989; 2: 51-78.

- Hamaker C. Item- en toetsindices. In: Creemers HPM, redactie. *Losbladig Onderwijskundig Lexicon*. Alphen aan den Rijn: Samsom, 1980.
- Hambleton RK. Criterion-referenced measurement. In: Keeves JP, editor. *Educational Research, Methodology and Measurement. An International Handbook*. Oxford: Pergamon Press, 1988.
- Harden RM. Constructing multiple choice questions of the multiple true/false type [medical education booklet nr. 10]. Dundee: ASME, 1979.
- Harden RM, Laidlaw JM. Effective continuing education: the CRISIS criteria. *Med Educ* 1992; 26: 408-22.
- Harteloh PPM, Casparie AF. *Kwaliteit van zorg. Van een zorginhoudelijke benadering naar een bedrijfskundige aanpak*. 's Gravenhage/Lochem: VUGA/De Tijdstroom, 1991.
- Haynes RB, Davis DA, McKibbon A, Tugwell P. A critical appraisal of the efficacy of continuing medical education. *JAMA* 1984; 251: 61-4.
- Hays RB, Bridges-Webb C, Booth B. Quality assurance in general practice. *Med Educ* 1993; 27: 175-80.
- Hays R. Methods of assessment in recertification. In: Newble D, Jolly B, Wakeford R, editors. *The Certification and Recertification of Doctors. Issues in the Assessment of Clinical Competence*. Cambridge: Cambridge University Press, 1994.
- Hessen P van, Imbos Tj. RIT's en DI's in de voortgangstoets [intern rapport]. Maastricht: PES-publ. 183, 1987.
- Hessen P van, Verwijnen M. Necessity of a test review committee in test construction. Paper presented at 'The International Symposium on Evaluation in Medical Education'; 1987 May 25-28; Beer Sheva, Israel. Maastricht: PES-158, 1987.
- Hessen PAW van, Verwijnen GM, Imbos Tj. De kennis van de Nederlandse basisartsen gemeten met de Maastrichtse voortgangstoets. *Ned Tijdschr Geneesk* 1991; 135: 1975-8.
- Hingstman L, Pool JB. Cijfers uit de registratie van huisartsen, peiling 1992. Utrecht: NIVEL, 1992.
- Hingstman L, Pool JB. Cijfers uit de registratie van beroepen in de eerstelijnsgezondheidszorg 1992. Statistische gegevens per 1 januari 1992 over huisartsen en verloskundigen. Utrecht: NIVEL, 1992.
- Hobus P. Expertise van huisartsen. *Praktijkervaring, kennis en diagnostische hypothesevorming* [Dissertatie]. Maastricht: Rijksuniversiteit Limburg, 1994.
- Hobus PPM, Hofstra ML, Boshuizen HPA, Schmidt HG. De context van de klacht als diagnosticum. *Huisarts Wet* 1988; 31: 261-3.
- Hoeben WThJG. Onderwijs: evaluatie en effectiviteit. Een nabeschouwing. In: Hofstee WKB, redactie. *Evaluatiemethodologie*. Lisse: Swets & Zeitlinger, 1987.
- Hofstee WKB. Het causaliteitsvraagstuk in evaluatieonderzoek. In: Hofstee WKB, redactie. *Evaluatiemethodologie*. Lisse: Swets & Zeitlinger, 1987.
- Horder J, Bosanquet N, Stocking B. Ways of influencing the behaviour of general practitioners. *J R Coll Gen Pract* 1986; 36: 517-21.
- Imbos Tj. Het gebruik van einddoeltoetsen bij aanvang van de studie [Dissertatie]. Maastricht: Rijksuniversiteit Limburg, 1989.
- Jansen JJM, Tan LHC, Vleuten CPM van der, Grol RPTM. Toetsing van medisch-technische vaardigheden bij huisartsen: een vergelijking tussen drie verschillende methoden. In: Metz JCM, Scherpier AJJA, Houtkoop E, redactie. *Gezond Onderwijs-II*. Nijmegen: Universitair Publikatiebureau KUN, 1993.

- Jansen JJM, Tan LHC, Vleuten CPM van der, Pielage JC, Grol RPTM. Assessment of technical skills in general practice: comparing different formats for use in quality control. In: Harden RM, Hart IR, Mulholland H, editors. *Approaches to the Assessment of Clinical Competence*. Dundee: Centre for Medical Education, 1992.
- Jarjoura D, Brennan RL. A variance components model for measurement procedures associated with a table of specifications. *Appl Psychol Meas* 1982; 6: 161-71.
- Johnson EJ. Expertise and decision under uncertainty: performance and process. In: Chi MTH, Glaser R, Farr MJ, editors. *The Nature of Expertise*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1988.
- Jones TV, Gerrity MS, Earp J. Written case simulations: do they predict physicians' behavior? *J Clin Epidemiol* 1990; 43: 805-15.
- Kane MT. The validity of licensure examinations. *Am Psychol* 1982; 37: 911-8.
- Kane MT. The future of testing for licensure and certification examinations. In: Plake BS, Witt JC, editors. *The Future of Testing*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1986.
- Kane MT. Is predictive validity the gold standard or is it the holy grail of examinations in the professions? American Educational Research Association: Washington DC, 1987.
- Kane MT. An argument-based approach to validity. *Psychol Bull* 1992; 112: 527-35.
- Kane MT. The assessment of professional competence. *Eval Hlth Prof* 1992; 15: 163-82.
- Kane MT, Brennan RL. The generalizability of class means. *Rev Educ Res* 1977; 47: 267-92.
- Kanouse DE, Jacoby I. When does information change practitioners' behavior? *Int J Technol Assess Health Care* 1988; 4: 27-33.
- Kempen BM, Leeuwen R van, Bock GH de. Evaluatie van het 'SOA-spreekuur': het rendement van een nascholingscursus voor huisartsen. In: Vleuten CPM van der, Scherpbier AJJA, Pollemans MC, redactie. *Gezond Onderwijs-I*. Houten/Zaventem: Bohn, Stafleu, Van Loghum, 1992.
- KNMG Commissie Kwaliteit. Kwaliteitsborging door herregistratie van medische beroepsbeoefenaren. Voorstellen ter ontwikkeling van een integraal model voor herregistratie van specialisten, huisartsen, sociaal-geneeskundigen en verpleeghuisartsen. Utrecht: KNMG, 1990.
- Kosecoff J, Kanouse DE, Rogers WH, McCloskey L, Winslow CM, Brook RH. Effects of the National Institutes of Health Consensus Development Program on physician practice. *JAMA* 1987; 258: 2708-13.
- Kramer AWM, Pollemans MC. Nation-wide progress tests assessing knowledge in vocational training for general practice. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, editors. *Teaching and Assessing Clinical Competence*. Groningen: BoekWerk Publ, 1990.
- Kramer AWM, Pollemans MC, Leeuwen YD van. Het toetsen van kennis in de beroepsopleiding tot huisarts. *Medisch Contact* 1990; 45: 220-2.
- Lamberts H. In het huis van de huisarts. Verslag van het Transitieproject. Lelystad: Meditekst, 1991.
- Lamberts H, Wood M, editors. *ICPC. International Classification of Primary Care*. Oxford/New York/Toronto: Oxford University Press, 1987.
- Langsley DG. Recredentialing [editorial]. *JAMA* 1991; 265: 772.
- Lawrence M, Pritchard P, editors. *General Practice Education. UK and Nordic perspectives*. London: Springer-Verlag, 1992.
- Laxdal OE. Needs assessment in continuing medical education: a practical guide. *J Med Educ* 1982; 57: 827-34.

- Leeuwen YD van. [Dissertatie]. Maastricht: Rijksuniversiteit Limburg, (in voorbereiding).
- Leeuwen YD van, Hessen PAW van. Clinical competence and objective questions. Tactics to realize a true/false format assessing competence. In: Bender W, Hiemstra RJ, Scherpbier AJA, Zwierstra RP, editors. Teaching and Assessing Clinical Competence. Groningen: BoekWerk Publ, 1990.
- Leeuwen YD van, Mol SSL, Pollemans MC, Drop MJ, Grol R, Vleuten CPM van der. Change in knowledge of general practitioners during their professional career. (in voorbereiding).
- Leeuwen YD van, Pollemans MC, Düsman H, Vleuten CPM van der, Grol RPTM. De huisartsgeneeskundige kennistoets. Constructvaliditeit en betrouwbaarheid. Welke maten meten wat? In: Metz JCM, Scherpbier AJA, Houtkoop E. Gezond Onderwijs-II. Nijmegen: Universitair Publikatiebureau KUN, 1993.
- Leeuwen YD van, Pollemans MC, Düsman H, Vleuten CPM van der, Grol R. Reliability of a case-based knowledge test for general practice trainees. (in voorbereiding).
- Leeuwen YD van, Pollemans MC, Eekhof JAH, Mol SSL. De ontwikkeling van de huisartsgeneeskundige kennistoets: Vraagvorm en blauwdruk. In: Vleuten CPM van der, Scherpbier AJA, Pollemans MC, editors. Gezond Onderwijs-I. Houten/Zaventem: Bohn, Stafleu, van Loghum, 1992.
- Leeuwen YD van, Pollemans MC, Mol SSL, Eekhof JAH, Grol R, Drop MJ. The Dutch knowledge test for general practice: issues of validity. (in voorbereiding).
- Leeuwen YD van, Pollemans MC, Verwijnen GM. Heiligt het doel de middelen? Toetsing met het oog op selectie en (her)registratie. Medisch Contact 1991; 46: 732-4.
- Leigh TM, Johnson TP, Pisacano NJ. Predictive validity of the American Board of Family Practice In-Training Examination. Acad Med 1990; 65: 454-7.
- Leinhardt G, Seewald AM. Overlap: what's tested, what's taught? J Educ Meas 1981; 18: 85-96.
- Levine HG, Cordes DL, Moore DE, Pennington FC. Identifying and assessing needs to relate continuing education to patient care. In: Green JS, Grosswald SJ, Suter E, Walthall DB, editors. Continuing Education for the Health Professions. San Francisco, etc.: Jossey-Bass Publ, 1984.
- Levine HG, McGuire CH, Nattress LW. The validity of multiple choice achievement tests as measures of competence in medicine. Am Educ Res J 1970; 7: 69-82.
- Levine HG, Moore DE, Pennington FC. Evaluating continuing education activities and outcomes. In: Green JS, Grosswald SJ, Suter E, Walthall DB, editors. Continuing Education for the Health Professions. San Francisco. etc.: Jossey-Bass Publ, 1984.
- LHV. De eerstelijnsgezondheidszorg en de positie van de huisarts in de komende jaren. Utrecht: Landelijke Huisartsen Vereniging, 1986.
- LHV. De positie van de huisarts in de toekomst [discussienota]. Utrecht: Landelijke Huisartsen Vereniging, 1987.
- LHV. Kwaliteit en deskundigheidsbevordering. Utrecht: Landelijke Huisartsen Vereniging, 1990.
- LHV. Modelcurriculum deskundigheidsbevordering huisartsen. Utrecht: Landelijke Huisartsenvereniging, 1993-I.
- LHV. Werkplan Deskundigheidsbevordering 1994. Utrecht: Landelijke Huisartsen Vereniging, 1993-II.
- Linden WJ van der. Het klassieke testmodel, latente trekmodellen en evaluatie-onderzoek. Enschede: Twents Onderwijskundig Memorandum, 1978.

- Linn RL. Change assessment. In: Keeves JP, editor. *Educational Research, Methodology and Measurement. An International Handbook*. Oxford: Pergamon Press, 1988.
- Lisdonk EH van de, redactie. *Ziekten in de huisartspraktijk*. Utrecht: Bunge, 1991.
- Lloyd JS, Abrahamson S. Effectiveness of Continuing Medical Education: a review of the evidence. *Eval Hlth Prof* 1979; 2: 251-80.
- Lockie C, editor. *Examination for Membership of the Royal College of General Practitioners (MRCGP). Development, Current State and Future Trends [Occasional Paper 46]*. London: Royal College of General Practitioners, 1990.
- Lockyer JM. Physician performance: the roles of knowledge, skills and environment. *Teach Learn Med* 1992; 4: 86-96.
- Lomas J, Anderson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med* 1989; 321: 1306-11.
- Lowenthal W. Continuing education for professionals. Voluntary or mandatory? *J Higher Educ* 1981; 51: 519-38.
- Mann KV, Chaytor KM. Help! Is anyone listening? An assessment of learning needs of practicing physicians. *Acad Med* 1992; 67: S4-S6.
- Manning PR, DeBakey L. Lifelong learning tailored to individual clinical practice. *JAMA* 1992; 268: 1135-6.
- Manning PR, Lee PV, Clinthworth WA, Denson TA, Oppenheimer PR, Gilman NJ. Changing prescribing practices through individual continuing education. *JAMA* 1986; 256: 230-2.
- McGuire C. Written methods for assessing clinical competence. In: Hart IR, Harden RM, editors. *Further Developments in Assessing Clinical Competence*. Montreal: Heal Publ, 1987.
- McLaughlin PJ, Donaldson JF. Evaluation of continuing medical education programs: a selected literature, 1984-1988. *J Contin Educ Health Prof* 1991; 11: 65-84.
- Means RP. How family physicians use information sources: implications for new approaches. In: Green JS, Grosswald SJ, Suter E, Walthall DB, editors. *Continuing Education for the Health Professions*. San Francisco, etc.: Jossey-Bass Publ, 1984.
- Meerling. *Methoden en technieken van psychologisch onderzoek* (2e druk). Meppel: Boom, 1981.
- Meskauskas JA, Webster GD. The American Board of Internal Medicine Recertification Examination: process and results. *Ann Intern Med* 1975; 82: 577-81.
- Messick S. The standard problem: meaning and values in measurement and evaluation. *Am Psychol* 1975; 30: 955-66.
- Messick S. Abilities and knowledge in educational achievement testing: the assessment of dynamic cognitive structure. In: Plake BS, editor. *Social and Technical Issues in Testing: Implications for Test Construction and Language*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1984.
- Messick S. The psychology of educational measurement. *J Educ Meas* 1984; 21: 215-37.
- Messick S. The once and future issues of validity: assessing the meaning and consequences of measurement. In: Wainer H, Braun HI, editors. *Test Validity*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1988.
- Messick S. Validity. In: Linn RL, editor. *Educational Measurement* (third edition). New York: American Council on Education/MacMillan Publishing Comp, 1989.
- Metz JCM. *Leren en onderwijzen van geneeskunde [oratie]*. Nijmegen: Katholieke Universiteit Nijmegen, 1993.

- Metz JCM, Pels Rijcken-van Erp Taalman Kip EH, Brand van den-Valkenburg BWM. Raamplan 1994 artsopleiding. Eindtermen van de artsopleiding. Nijmegen: Universitair Publikatiebureau KUN, 1994.
- Meijer LJ, Zwart Sj, Westerveld MC e.a. NHG-Standaard Vaginaal bloedverlies. Huisarts Wet 1992; 35: 475-81.
- Millac J. Continuing Medical Education. J R Coll Gen Pract 1985; 35: 303-4.
- Miller GE. Continuous assessment. Med Educ 1976; 10: 81-6.
- Miller GE. Continuing education. What it is and what it is not. JAMA 1987; 258: 1352-4.
- Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990; 65: S63-7.
- Millman J, Greene J. The specification and development of tests of achievement and ability. In: Linn RL, editor. Educational Measurement (third edition). New York: American Council on Education/Macmillan Publishing Company, 1989.
- Moelands H, Noyons J, Rem J. Toetsen met gesloten vragen. Een handleiding voor het construeren van toetsen met meerkeuzevragen. Arnhem: Cito, 1992.
- Mol SSL, Leeuwen YD van, Pollemans MC. An international perspective on assessment in general practice. In: Harden RM, Hart IR, Mulholland H, editors. Approaches to the Assessment of Clinical Competence. Dundee: Centre for Medical Education, 1992.
- Mol SSL, Pollemans MC, Leeuwen YD van, Vleuten CPM van der, Grol RPT. Het mondeling examen van het Britse college van huisartsen. 'Orals revisited'. Medisch Contact 1994; 49: 1091-3.
- Mueller CB. Implications for licensure and certification. In: Neufeld VR, Norman GR, editors. Assessing Clinical Competence. New York: Springer Publishing Company, 1985.
- Muir Gray J. Continuing Education: What techniques are effective? Lancet 1986; ii: 447-8.
- Mulholland H, Harden RM. Problems encountered in moving from norm-referenced to criterion-referenced assessment of clinical competence. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, editors. Teaching and Assessing Clinical Competence. Groningen: BoekWerk Publ, 1990.
- Neufeld VR. Written examinations. In: Neufeld VR, Norman GR, editors. Assessing Clinical Competence. New York: Springer Publishing Company, 1985.
- Neufeld VR, Norman GR, editors. Assessing Clinical Competence. New York: Springer Publishing Company, 1985.
- Newble DI. Assessing clinical competence at the undergraduate level [ASME Medical Education Booklet no. 25]. Med Educ 1992; 26: 504-11.
- Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. Med Educ 1983; 17: 165-71.
- Newble D, Jolly B, Wakeford R, editors. The Certification and Recertification of Doctors. Issues in the Assessment of Clinical Competence. Cambridge: Cambridge University Press, 1994.
- NHG. 'Naar criteria voor kwaliteit'. Standaardenbeleid NHG [discussienota]. Utrecht: Nederlands Huisartsen Genootschap, 1987.
- NHG. Werkboek Standaarden Ontwikkeling. Utrecht: Nederlands Huisartsen Genootschap, 1989.
- NHG. Jaarboek 1990-1991. Utrecht: Nederlands Huisartsen Genootschap, 1990.
- NHG. Jaarboek 1993-1994, waarin opgenomen de jaarverslagen 1990-1992. Utrecht: Nederlands Huisartsen Genootschap, 1993-I.

- NHG. Rapportage van de werkgroep toetsing. Utrecht: Nederlands Huisartsen Genootschap, 1993-II.
- NIVEL. Beroepen in de extramurale gezondheidszorg. Aanbod van beroepsbeoefenaren in de extramurale zorg. Utrecht: NIVEL, 1994.
- Nona DA, Kenny WR, Johnson DK. The effectiveness of continuing education as reflected in the literature of the health professionals. *Am J Pharm Ed* 1988; 52: 111-7.
- Norcini J, Dawson-Saunders B. Issues in recertification in North America. In: Newble D, Jolly B, Wakeford R, editors. *The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence*. Cambridge: Cambridge University Press, 1994.
- Norcini JJ, Lipner RS, Benson JA, Webster GD. An analysis of the knowledge base of practicing internists as measured by the 1980 recertification examination. *Ann Intern Med* 1985; 102: 385-9.
- Norcini JJ, Swanson DB. Factors influencing testing time requirements for measurements using written simulations. *Teach Learn Med* 1989; 1: 85-91.
- Norcini JJ, Swanson DB, Grosso LJ, Shea JA, Webster GD. A comparison of knowledge, synthesis, and clinical judgment. Multiple-choice questions in the assessment of physician competence. *Eval Hlth Prof* 1984; 7: 485-99.
- Norcini JJ, Swanson DB, Grosso LJ, Webster GD. Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. *Med Educ* 1985; 19: 238-47.
- Norman GR. Problem-solving skills, solving problems and problem-based learning. *Med Educ* 1988; 22: 279-86.
- Norman G. Can an examination predict competence? The role of recertification in maintenance of competence. *Annals RCPSC* 1991; 24: 121-4.
- Norman G, Allery L, Berkson L, et al. Research in the psychology of clinical reasoning: implications for assessment. In: Jolly B, editor. *New Directions in the Assessment of Clinical Competence*. Proceedings Cambridge Conference IV. Cambridge: Cambridge University Press, 1989.
- Norman GR, Rosenthal D, Brooks LR, Allen SW, Muzzin LJ. The development of expertise in dermatology. *Arch Dermatol* 1989; 125: 1063-8.
- Norman GR, Smith EKM, Powles ACP, Rooney PJ, Henry NL, Dodd PE. Factors underlying performance on written tests of knowledge. *Med Educ* 1987; 21: 297-304.
- Norman GR, Tugwell P, Feightner JW, Muzzin LJ, Jacoby LL. Knowledge and clinical problem-solving. *Med Educ* 1985; 19: 344-56.
- Norman GR, Vleuten CPM van der, Graaff E de. Pitfalls in the pursuit of objectivity II: issues of validity, efficiency and acceptability. *Med Educ* 1991; 25: 119-26.
- Oosterhof AC, Glasnapp DR. Comparative reliabilities and difficulties of the multiple-choice and true-false formats. *J Exp Educ* 1974; 42: 62-4.
- Palincsar AS, Winn J. Assessment models focused on new conceptions of achievement and reasoning. *Int J Educ Res* 1990; 14: 409-83.
- Parry KM. Effective continuing medical education. Report of a conference of the association for the study of medical education held at the National Heart and Lung Institute, London on 19 january 1990. *Med Educ* 1990; 24: 546-50.
- Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. *Cognit Sci* 1986; 10: 91-116.
- Patel VL, Groen GJ, Frederiksen CH. Differences between medical students and doctors in memory for clinical cases. *Med Educ* 1986; 20: 3-9.

- Pickering G. Against multiple choice questions. *Med Teach* 1979; 1: 84-6.
- Pickup AJ, Mee LG, Hedley AJ. Obstacles to continuing education. *J R Coll Gen Pract* 1983; 33: 799-801.
- Pietroni R. New strategies for higher professional education. *Br J Gen Pract* 1992; 42: 294-6.
- Pisacano NJ, Veloski JJ, Brucker PC, Gonnella JS. Classifying the content of board certification examinations. *Acad Med* 1989; ii: 149-54.
- Pollemans MC, Düsman H. Statistiek en meten: wat moet je daarover weten? Klassieke testtheorie en betrouwbaarheid van meetgegevens. Verschijnt in *Bulletin Medisch Onderwijs*, 1994.
- Pollemans MC, Eekhof JAH, Kramer AWM, Leeuwen YD van. Kennistoetsing bij aanstaande huisartsen. *Bulletin Medisch Onderwijs* 1991; 10: 76-80.
- Pollemans MC, Eekhof JAH, Verwijnen GM. Juist-onjuist vragen. In: Metz JCM, Scherpbier AJJA, Vleuten CPM van der, redactie. *Medisch Onderwijs in de Praktijk. Een gids voor docenten*. Assen: van Gorcum & Comp, (in voorbereiding).
- Pollemans MC, Geldorp G van, Tan LHC. Naar kwaliteitsbewaking van de beroepsopleiding tot huisarts. *Medisch Contact* 1988; 43: 1429-30.
- Pollemans MC, Leeuwen YD van, Düsman H, Eekhof JAH, Mol SSL. Achtergronden blauwdruk huisartsgeneeskundige kennistoets [rapport SV-IOH-21]. Utrecht: Samenwerkingsverband-IOH/WOK, 1993.
- Pollemans MC, Leeuwen YD van, Mol SSL, Düsman H. (Zelf)kennis van huisartsen. In: Metz JCM, Scherpbier AJJA, Houtkoop E, redactie. *Gezond Onderwijs-II*. Nijmegen: Universitair Publikatiebureau KUN, 1993.
- Pollemans MC, Leeuwen YD van, Mol SSL, Grol R. Toetsing van huisartsgeneeskundige kennis. Ontwikkeling en evaluatie van kennistoetsen voor opleiding en nascholng. Maastricht: Werkgroep Onderzoek Kwaliteit KUN-RL (WOK), Utrecht: Samenwerkingsverband Universitaire Huisartsopleidingen (SVUH)/Nederlands Huisartsen Genootschap (NHG), 1994.
- Pollemans MC, Tan LHC. Samenwerking in de beroepsopleiding tot huisarts. *Bulletin Medisch Onderwijs* 1988; 7: 23-7.
- Pollemans MC, Tan LHC. Toetsing van kwaliteit. Landelijke evaluatie van de interim-beroepsopleiding tot huisarts [rapport SV-IOH-15]. Utrecht: Samenwerkingsverband-IOH, 1990.
- Polliack MR. Some ethical issues in Continuing Medical Education. *Med Law* 1990; 9: 791-5.
- Prawat RS. Promoting access to knowledge, strategy and disposition in students: a research synthesis. *Rev Educ Res* 1989; 20: 1-41.
- Ramsey PG, Carline JD, Inui TS, Larson EB, LoGerfo JP, Wenrich MD. Predictive validity of certification by the American Board of Internal Medicine. *Ann Int Med* 1989; 110: 719-26.
- Relman AS. Assessment and accountability: the third revolution in medical care. *N Engl J Med* 1988; 319: 1220-3.
- Rethans JJ. Does competence predict performance? Standardized patients as a means to investigate the relationship between competence and performance of general practitioners [Dissertatie]. Amsterdam: Thesis Publ, 1991.
- RGO-werkgroep Kwaliteit van zorg. Advies Kwaliteit van zorg. Terreinverkenning en prioriteiten voor wetenschappelijk onderzoek [advies no. 6]. 's Gravenhage: Raad voor Gezondheidsonderzoek, 1990.

- Richardson R. The multiple choice true/false question: what does it measure and what could it measure? *Med Teach* 1992; 14: 201-4.
- Rosinski EF. The relationship between continuing medical education, physician behavior and patient outcomes. *Möbius* 1987; 7: 16-21.
- Rutten GEH, Thomas S, redactie. *NHG-Standaarden voor de huisarts*. Utrecht: Bunge/NHG, 1993.
- Sackett DL, Haynes RB, Taylor DW et al. Process versus outcome in the clinical management of hypertension. *Clin Res* 1979; 27: 443A.
- Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology. A basic science for medicine*. Boston/Toronto: Little, Brown and Company, 1985.
- Salem-Schatz SR, Avorn J, Soumerai SB. Influence of clinical knowledge, organizational context, and practice style on transfusion decision making. *JAMA* 1990; 264: 476-83.
- Sanazaro PJ. Determining physicians' performance. *Continuing medical education and other interacting variables*. *Eval Hlth Prof* 1983; 6: 197-210.
- Schadé E, Lamberts H. Gebruik van de ICPC in de huisartsgeneeskunde. In: Hirs WM, Flier F, Halbertsma J, Planten AThH, redactie. *Classificatie en terminologie in de gezondheidszorg (verslag WCC-conferentie)*. Bussum, 1989.
- Schmidt HG, Boshuizen HPA. On acquiring expertise in medicine. *Educational Psychology Review* 1993; 5: 205-21.
- Schmidt HG, Boshuizen HPA, Hobus PPM. Transitory stages in the development of medical expertise: the 'intermediate effect' in clinical case representation studies. In: *Proceedings of the 10th Annual Conference of the Cognitive Science Society*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1988.
- Schmidt HG, Boshuizen HPA, Norman GR. Reflections on the nature of expertise in medicine. In: Keravnou E, editor. *Deep Models for Medical Knowledge Engineering*. Amsterdam: Elsevier Science Publ., 1992.
- Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory and implications. *Acad Med* 1990; 65: 611-21.
- Schofield TPC. Continued medical education must not be an optional extra. *Br Med J* 1987; 294: 526-7.
- Schuwirth LWT, Jean P, Vleuten CPM van der, Santen M van. *Problem Analysis Questions, een korte casusvorm voor het pre-klinisch domein*. Bijdrage congres Gezond Onderwijs-III. Veldhoven, 1993.
- Schuwirth LWT, Schrandt JPP, Vleuten CPM van der. Assistententoets kindergeneeskunde: een beschrijving van de psychometrische eigenschappen. *Bulletin Medisch Onderwijs* 1993; 12: 146-51.
- SDH. *Beleidsplan en structuur van de deskundigheidsbevordering voor huisartsen 1991-1995*. Utrecht: Stichting Deskundigheidsbevordering Huisartsen, 1990.
- Sibley JC, Sackett DL, Neufeld V, Gerrard B, Rudnick KV, Fraser W. A randomized trial of continuing medical education. *N Engl J Med* 1982; 306: 511-5.
- Simons PRJ. *Transfervermogen [oratie]*. Nijmegen: Katholieke Universiteit Nijmegen, 1990.
- Simons PRJ, Verschaffel L. *Transfer: onderzoek en onderwijs*. Tijdschrift voor Onderwijsresearch 1992; 17: 3-16.
- Slotboom A. *Statistiek in woorden*. Groningen: Wolters-Noordhoff, 1987.
- Sluijs EM, Bakker DH de. *Kwaliteitssystemen in ontwikkeling*. Utrecht: NIVEL, 1992.
- Snoek JW. *Het denken van de neuroloog [Dissertatie]*. Groningen: Rijksuniversiteit Groningen, 1989.

- Soeters D, Scherpbier AJJA, Vleuten CPM van der, Bootsma-de Langen AM. Het effect van een nascholingscursus voor huisartsen. Is it all worth-while or just for fun? In: Metz JCM, Scherpbier AJJA, Houtkoop E. Gezond Onderwijs-II. Nijmegen: Universitair Publikatiebureau KUN, 1993.
- Southgate L, Jolly B. Determining the content of recertification procedures. In: The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence. Cambridge: Cambridge University Press, 1994.
- Spike NA, Veith PC. Analysis of the RACGP Fellowship examination results. *Aus Fam Physician* 1990; 19: 767-75.
- Spiro RJ, Vispoel WP, Schmitz JG, Samarapungavan A, Boerger AE. Knowledge acquisition for application: cognitive flexibility and transfer in complex content domains. In: Britton BK, Glynn SM, editors. Executive control processes in reading. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1987.
- Springer MP, redactie. Basistakenpakket voor de huisarts. Utrecht: Landelijke Huisartsen Vereniging, 1983.
- Stanley J, Al-Shehri A. Reaccreditation: the why, what and how questions. *Br J Gen Pract* 1993; 43: 524-9.
- Stanley J, Al-Shehri A, Thomas P. Continuing education for general practice. 1. Experience, competence and the media of self-directed learning for established general practitioners. *Br J Gen Pract* 1993; 43: 210-4.
- Steel K, Norcini J, Brummel-Smith K, Erwin D, Markson L. The first certifying examination in geriatric medicine. *J Am Geriatr Soc* 1989; 37: 1188-91.
- Stein LS. The effectiveness of continuing medical education: eight research reports. *J Med Educ* 1981; 56: 103-10.
- Suen HK. Principles of Test Theories. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1990.
- Susteren TJ van, Cohen EB, Simpson DE. Alternate-choice test items: implications for measuring clinical judgment. *Teach Learn Med* 1991; 3: 33-7.
- Suter E, Green JS, Grosswald SJ, Lawrence KA, Walthall DB, Zeleznik C. Introduction: defining quality for continuing education. In: Green JS, Grosswald SJ, Suter E, Walthall DB, editors. Continuing Education for the Health Professions. San Francisco, etc.: Jossey-Bass Publ, 1984.
- Suter E, Green JS, Lawrence K, Walthall DB. Continuing education of health professionals: proposal for a definition of quality. *J Med Educ* 1981; 56: 689-707.
- Swanson D, Case S. Trends in written assessment: a strangely biased perspective. In: Harden RM, Hart IR, Mulholland H, editors. Approaches to the Assessment of Clinical Competence. Dundee: Centre for Medical Education, 1992.
- Swanson DB, Case SM, Vleuten CPM van der. Strategies for student assessment. In: Boyd D, Feletti G, editors. The Challenge of Problem-based Learning. London: Kogan Page, 1991.
- Swanson DB, Norcini JJ, Grosso LJ. Assessment of clinical competence: written and computer-based simulations. *Assessment and Evaluation in Higher Education* 1987; 12: 220-46.
- Tan LHC. Tekorten in de opleiding van huisartsen. Ziektebeelden en medisch-technische vaardigheden [Dissertatie]. Amsterdam: Universiteit van Amsterdam, 1989.

- Tan LHC, Jansen JJM, Pielage JC. Toetsing van medisch-technische vaardigheden in de huisartsgeneeskunde. De SV-WOK Vaardigheidentoets maart 1992 [rapport SV-IOH-23]. Utrecht: Samenwerkingsverband-IOH/WOK, 1993.
- Tan LHC, Pollemans MC. Verslag project toetselementen productie 1991 (TEP-91) [rapport SV-IOH-20]. Utrecht: Samenwerkingsverband-IOH, 1992.
- Thorndike RL, editor. Educational Measurement (second edition). Washington DC: American Council on Education, 1971.
- Thorndike RL, editor. Applied Psychometrics. Boston, etc.: Houghton Mifflin Company, 1982.
- Verwijnen GM, Imbos Tj, Hessen PAW van, Wijnen WHFW. What's in the score! An exploratory analysis of true/false test scores. Paper presented at 'The International Symposium on Evaluation in Medical Education'; 1987 May 25-28; Beer Sheva, Israel, 1987.
- Verwijnen GM, Imbos Tj, Snellen H. et al. The evaluation system at the Medical School of Maastricht. Assessment and Evaluation in Higher Education 1982; 7: 225-44.
- Verwijnen GM, Pollemans MC, Wijnen WHFW. Voortgangstoetsing. In: Metz JCM, Scherpbier AJJA, Vleuten CPM van der, redactie. Medisch Onderwijs in de Praktijk. Een gids voor docenten. Assen: van Gorcum & Comp, (in voorbereiding).
- Vleuten CPM van der. MGENOVA: a computer program for multivariate generalizability analysis for tests using table of specifications [intern rapport]. Maastricht: University of Limburg, 1991.
- Vleuten CPM van der. Toetsing van medische competentie: stand van zaken en ontwikkelingen. Bulletin Medisch Onderwijs 1991; 10: 52-63.
- Vleuten CPM van der, Luyk SJ van, Beckers HJM. A written test as an alternative to performance testing. Med Educ 1989; 23: 97-107.
- Vleuten C van der, Newble D. Methods of assessment in certification. In: Newble D, Jolly B, Wakeford R, editors. The Certification and Recertification of Doctors: Issues in the Assessment of Competence. Cambridge: Cambridge University Press, 1994.
- Vleuten CPM van der, Norman GR. Implications of research on licensure and certification for recertification. Position paper for Cambridge Conference, University of Adelaide, 1991.
- Vleuten CPM van der, Norman GR, Graaff E de. Pitfalls in the pursuit of objectivity I: issues of reliability. Med Educ 1991; 25: 110-8.
- Vleuten CPM van der, Verwijnen M. A system for student assessment. In: Vleuten C van der, Wijnen W, editors. Problem-based learning: Perspectives from the Maastricht experience. Amsterdam: Thesis, 1990.
- Vleuten CPM van der, Wijnen WHFW. Niets praktischer dan een goede theorie: generaliseerbaarheidstheorie als instrument voor betrouwbaarheidsstudies. Bulletin Medisch Onderwijs 1991; 10: 2-14.
- Vleuten M van der. Voortgangstoetsen: het verloop van de betrouwbaarheid bij verandering van het aantal vragen [intern rapport]. Maastricht: PES-publ. 97, 1985.
- Waldrop MM. The necessity of knowledge. Science 1984; 223: 1279-82.
- Weiss CH. Evaluation Research. Methods of assessing program effectiveness. Englewood Cliffs, New Jersey: Prentice-Hall Inc, 1972.
- Wergin JF, Mazmanian PE, Miller WW, Papp KK, Williams WL. Assessing the impact of continuing medical education through structured physician diagnosis. Bethesda: American College of Cardiology, 1986.
- Weijden T van der. Implementatieproject Cholesterol-Standaard. Maastricht: Vakgroep Huisartsgeneeskunde, 1992.

- Weijden T van der, Grol RPTM, Schouten BJ, Knottnerus JA. Implementation of guidelines on cholesterol in general practice. A randomized controlled trial. (in voorbereiding).
- Wijnen WHFW. Kwaliteitsbewaking met voortgangstoetsen [intern rapport]. Maastricht: PES-publ. 119, 1986.
- Wijnen WHFW. Voordracht tijdens NVMO-studiedag over toetsing. Nijmegen, 9 juni 1993.
- Zuidweg J. Diabeteskennistoetsen in de beroepsopleiding tot huisarts [Dissertatie]. Maastricht: Rijksuniversiteit Limburg, 1994.

Wetenschappelijke kwaliteitsbewaking

BIJLAGEN

1. Vragenlijst betreffende de inhoud en structuur van de huisartsgeneeskundige kennistoets
2. Definities nevenindelingen blauwdruk
3. Generaliseerbaarheidstheorie
4. Formules betrouwbaarheidsstudies
5. Voorbeeldpagina algemene huisartsgeneeskundige kennistoets februari 1992
6. Vragenlijst achtergrondkenmerken
7. Satisfactievragenlijst landelijke kennistoets februari 1992
8. Instructiebrief huisartsen-deelnemers landelijke kennistoets februari 1992
9. Voorbeeld individuele feedback kennistoets februari 1992, plus toelichtende brief
10. Voorbeeldpagina kennistoets vaginaal bloedverlies
11. Inventarisatie kenniselementen NHG-Standaard Cholesterol
12. Voorbeeldpagina kennistoets cholesterol
13. Satisfactievragenlijst onderwerpgebonden toetsen
14. Instructiebrief huisartsen-deelnemers onderwerpgebonden toetsing
15. Voorbeeld individuele feedback onderwerpgebonden toetsing, plus toelichtende brief
16. Retentievragenlijst kennistoets over cholesterol
17. Voorbeeldpagina kennistoets cholesterol (meerkeuzevragen)
18. Voorbeeld individuele feedback meerkeuzetoets, plus toelichtende brief

BIJLAGE 1 VRAGENLIJST BETREFFENDE DE INHOUD EN STRUCTUUR VAN DE HUISARTSGENEESKUNDIGE KENNISTOETS

Het doel van de vragenlijst

In de vragenlijst worden de meest cruciale keuzes ten aanzien van een voorstel voor de structuur en de inhoud van relevante huisartsgeneeskundige kennistoetsen toegelicht. De vragenlijst dient als onderlegger voor groepsdiscussies over de keuzes en de argumenten die aan het voorstel ten grondslag liggen, met vertegenwoordigers van het huisartsgeneeskundig beroepsveld, inclusief de beroepsopleiding. Doel van deze groepsdiscussies is het bereiken van overeenstemming over een goede structuur voor de huisartsgeneeskundige kennistoets.

Toelichting op de noodzaak van een vaste structuur voor de kennistoetsen

De kwaliteit van de kennistoets als instrument om het kennisniveau van huisartsen(-in-opleiding) en veranderingen (eventueel groei) daarin te kunnen vaststellen, wordt bepaald door inhoudelijke en psychometrische (toetstechnische) factoren. Zo moet een kennistoets bijvoorbeeld bestaan uit een voldoende grote verzameling toetsvragen, die allemaal huisartsgeneeskundig relevant, qua formulering niet voor misverstand vatbaar en vakinhoudelijk onderbouwd zijn. Ook moet elke toets volgens dezelfde vaste indeling (blauwdruk) zijn samengesteld om te kunnen waarborgen, dat elke toets een representatieve steekproef van vragen uit het gehele domein van de huisartsgeneeskunde bevat.

Deze **standaard-indelingsstructuur** met vaste hoofdstukken en vaste aantallen toetsvragen per hoofdstuk is noodzakelijk om ervoor te kunnen zorgen dat verschillende toetsen 'het zelfde' meten. Men spreekt dan van toetsen die inhoudelijk parallel zijn. Inhoudelijke paralleliteit van de kennistoetsen is een van de voorwaarden om de resultaten (zowel van één bepaalde persoon als van de hele groep) op opeenvolgende toetsen met elkaar te mogen vergelijken.

De kennistoets zoals die tot nu toe, experimenteel, in de beroepsopleiding wordt afgenomen, kent nog niet zo'n vaste indelingsstructuur. Elke nieuwe toets wordt samengesteld uit acht willekeurige onderwerpen (zie bijlage). Deze onderwerpen worden gekozen door de auteurs van vragen. Deze auteurs zijn telkens andere stafleden van de acht Nederlandse huisartsinstituten.

Het onderzoek naar de kwaliteit van de huisartsgeneeskundige kennistoets, is gestart met het formuleren van een voorstel voor een standaard-indelingsstructuur voor de kennistoetsen. De vragen in de hierna volgende paragrafen hebben betrekking op de keuzes die aan dat voorstel ten grondslag liggen.

De onderwerpen van de vragenlijst

De vragen die wij u hierna voorleggen hebben betrekking op de volgende onderwerpen:

- de (hoofd)indeling van de toets in hoofdstukken;
- de groepering van deze hoofdstukken tot clusters, waarbij wordt betrokken de meest haalbare toetslengte;
- de (neven)indeling van de toets volgens andere dimensies.

Wij willen u vragen om per paragraaf allereerst de toelichting op het geformuleerde voorstel goed door te lezen, voordat u de bijbehorende vraag of vragen beantwoordt.

1. De hoofdingeling van de kennistoets

De eerste vraag die wij u voorleggen behelst de hoofdingeling van de toets. De vraag luidt globaal: uit welke hoofdstukken dient de toets te bestaan?

Door ons werd een aantal criteria geformuleerd waaraan de hoofdingeling zou moeten voldoen. Deze luiden als volgt.

De hoofdingeling van de toets dient:

- de **dekking** van het gehele gebied van de huisartsgeneeskunde, dat voor kennistoetsing in aanmerking komt, te kunnen garanderen;
- zo breed te zijn dat **méér dan alleen het klinisch handelen en zeker meer dan alleen het somatisch klinisch handelen** kan worden omvat;
- **herkenbaar** te zijn voor de huisarts, met ander woorden: de hoofdonderwerpen moeten een voorstelbare inhoud hebben voor de huisarts;
- zoveel mogelijk **aan te sluiten bij de wijze waarop** in de dagelijkse praktijk en daarbuiten kennis wordt verworven;
- liefst **aan te sluiten bij andere indelingen** (bijvoorbeeld bij indelingen die worden gebruikt in andere toetsen of bij morbiditeitsregistraties);
- uit hoofdstukken te bestaan, waarover een voldoende aantal vragen geformuleerd en opgenomen kan worden.

Uit de literatuur blijkt dat de meest omvattende omschrijving op huisartsgeneeskundig gebied wordt geleverd door de **International Classification of Primary Care (ICPC)**. De ICPC wordt op Europese schaal gebruikt voor het classificeren van het morbiditeitsaanbod in de huisartspraktijk en is opgebouwd uit rubrieken die elkaar onderling niet overlappen. Het gebruik van de ICPC als basis voor de indeling van de kennistoets in hoofdstukken, betekent dat aan de meeste van de genoemde criteria kan worden voldaan: de ICPC is herkenbaar voor de beroepsgroep, laat toe verbanden te leggen tussen morbiditeit, onderwijs en toetsgegevens en omvat een medisch-inhoudelijk breed spectrum. De ICPC voorziet echter **niet** in een huisartsgeneeskundig-theoretische component met onderwerpen die met name voor opleiding en scholing van belang zijn (zoals kennis over methoden van wetenschappelijk onderzoek, beslissonderzoek, gezinsgeneeskunde, somatische fixatie etc.). Dit zou kunnen worden ondervangen door het toevoegen van een extra hoofdstuk.

Als mogelijk alternatief werd overwogen een indeling gebaseerd op medische vakgebieden ofwel disciplines (interne, chirurgie enz.). De nadelen van een dergelijke indeling leken reeds op voorhand te groot: de vakken overlappen elkaar en de indeling zou te medisch-specialistisch van aard worden.

Voorgesteld wordt derhalve om in de blauwdruk de ICPC als **hoofdingelingsprincipe** voor de kennistoets te hanteren. Het voorstel behelst bovendien de opname van een huisartsgeneeskundig-theoretisch hoofdstuk naast de ICPC-hoofdstukken. Het voorstel voor de hoofdingeling van de kennistoets ziet er als volgt uit.

Voorstel hoofdingeling huisartsgeneeskundige kennistoets

1. Algemeen en niet gespecificeerd (ICPC-hoofdstuk A)
2. Bloed en bloedvormende organen (ICPC-hoofdstuk B)
3. Tractus digestivus (ICPC-hoofdstuk D)
4. Oog (ICPC-hoofdstuk F)
5. Oor (ICPC-hoofdstuk H)
6. Tractus circulatorius (ICPC-hoofdstuk K)
7. Bewegingsapparaat (ICPC-hoofdstuk L)
8. Zenuwstelsel (ICPC-hoofdstuk N)
9. Psychische problemen (ICPC-hoofdstuk P)
10. Tractus respiratorius (ICPC-hoofdstuk R)
11. Huid en subcutis (ICPC-hoofdstuk S)
12. Endocriene klieren, metabolisme, voeding (ICPC-hoofdstuk T)
13. Urinewegen (ICPC-hoofdstuk U)

14. Zwangerschap, bevalling, anticonceptie (ICPC-hoofdstuk W)
15. Geslachtsorganen en borsten vrouw (ICPC-hoofdstuk X)
16. Geslachtsorganen en borsten man (ICPC-hoofdstuk Y)
17. Sociale problemen (ICPC-hoofdstuk Z)
18. Huisartsgeneeskundig-theoretische onderwerpen (nieuw: b.v. beslistkunde, methodologie, tuchtrecht, kennis over ziektewet)

VRAAG 1A: Bent u het eens met de keuze voor de ICPC als uitgangspunt voor de hoofdingeling voor de huisartsgeneeskundige kennistoetsen?
Zo nee: welk alternatief stelt u voor?

VRAAG 1B: Bent u het eens met de voorgestelde toevoeging van de extra categorie huisartsgeneeskundig-theoretische onderwerpen?
Zo nee: welk alternatief stelt u voor?

VRAAG 1C: Mist u relevante onderwerpen?

2. Clustering van de hoofdstukken

Wij leggen u nu de vraag voor op welke wijze de hoofdstukken uit de hoofdingeling het beste zouden kunnen worden gegroepeerd, en wel op zodanige wijze dat maximaal tien clusters ontstaan die elk op zich weer begrijpelijke en zinvolle eenheden vormen. De noodzaak van een dergelijke clustering lichten we hierna toe.

In de standaard-indelingsstructuur (blauwdruk) van de kennistoets dient ook de lengte van de toets, c.q. het aantal toetsvragen, te zijn omschreven. Enerzijds gaat het om een **minimum**lengte, c.q. minimum aantal toetsvragen. Dit aantal is afhankelijk van de eisen waaraan de toets moet voldoen om van een aanvaardbare **betrouwbaarheid** van de toetsresultaten te kunnen spreken. Anderzijds gaat het om een **maximum**lengte, c.q. maximum aantal toetsvragen. Dit aantal is afhankelijk van de voorwaarden die gelden voor de **acceptabiliteit** van de toets: de toets moet niet al op voorhand door zijn lengte ongenietbaar worden.

De huidige huisartsgeneeskundige kennistoets bevat gemiddeld 160 vragen, dat wil zeggen 20 vragen per onderwerp. (Een aantal van 20 lijkt het absolute minimum om over het resultaat op onderdelen van de toets nog uitspraken te mogen doen.) De tijd die nodig is om deze toets te beantwoorden is gemiddeld 2 uur. De kennistoets zal in totaal waarschijnlijk iets méér vragen moeten bevatten om tot een aanvaardbare waarde voor de betrouwbaarheid van de toets als geheel te komen. Een aantal van meer dan 200 vragen is waarschijnlijk niet meer 'genietbaar'.

De voorgestelde hoofdingeling gebaseerd op de ICPC telt 18 hoofdstukken. 18 maal 20 = 360 vragen. Dit aantal lijkt de grenzen van het aanvaardbare (geschatte invulduur 4,5 uur) ruimschoots te overschrijden. Bovendien zouden er problemen ontstaan in verband met het formuleren van de benodigde aantallen toetsvragen. Samenvoegen van ICPC-hoofdstukken tot clusters (toetsonderwerpen) is daarom noodzakelijk.

Hieronder vindt u een voorstel tot samenvoegen van ICPC-hoofdstukken. De keuze van de clusters is deels voor de hand liggend (b.v. urogenitaal stelsel van man en vrouw), deels arbitrair (b.v. zenuwen en zintuigen). De in het voorstel opgenomen percentages representeren het aandeel van de onderdelen van elk cluster in de toets. Deze percentages zijn onder meer gebaseerd op **morbiditeitscijfers**. Uitgangspunt was tevens dat de verschillende clusters een ongeveer gelijk aandeel in de toets moesten

innemen om ervoor te kunnen zorgen dat de toetsresultaten per cluster qua betrouwbaarheid onderling vergelijkbaar zouden zijn.

Voorstel tot clustering van de ICPC-hoofdstukken

I	- algemeen (A), psych. (P), sociale problemen (Z):	11%
	(A: 5% + PZ: 6%)	
II	- tr. digestivus (D), voeding/endoctr./metabool (T):	11%
	(D: 7% + T: 4%)	
III	- oog (F), oor (H), zenuwstelsel (N):	11%
	(F: 3,5% + H: 3,5% + N: 3,5%)	
IV	- tr. circulatorius (K), bloed (B):	11%
	(K: 10% + B: 1,5%)	
V	- bewegingsapparaat (L):	11%
	(L: 11%)	
VI	- tr. respiratorius (R):	11%
	(R: 11%)	
VII	- huid en subcutis (S):	11%
	(S: 11%)	
VIII	- tr. urogenitalis (U + W + X + Y):	11%
	(U: 2,5% + W: 3% + X: 3% + Y: 2,5%)	
IX	- huisartsgeneeskundig-theoretisch:	12%
totaal:		100%

VRAAG 2A: Bent u het eens met de hier voorgestelde clustering van ICPC-hoofdstukken uit de hoofdingeling?

Zo nee, wat is uw alternatief?

VRAAG 2B: Bent u het eens met morbiditeitscijfers als uitgangspunt voor de verdeling van het aantal vragen over de verschillende onderwerpen?

Zo nee, wat is uw alternatief?

VRAAG 2C: Wat is volgens u de maximaal acceptabele invultijd voor de huisartsgeneeskundige kennistoets?

3. De nevenindeling van de huisartsgeneeskundige kennistoets

Behalve de vakinhoudelijke 'dekking' van het gehele huisartsgeneeskundige kennisgebied, die door de ICPC als hoofdingelingsprincipe wordt gegarandeerd, bestaan er nog meer relevante dimensies die bij het samenstellen van goede huisartsgeneeskundige kennistoetsen een rol zouden moeten spelen. Zo kan er bijvoorbeeld voor worden gewaakt dat, over de gehele toets genomen, alle leeftijdscategorieën van patiënten voldoende aan bod komen of dat de toets een evenredig deel acute en chronische problemen bevat. Het betreft hier zogenaamde **nevenindelingen** voor de huisartsgeneeskundige kennistoets.

Onze derde vraag betreft deze nevenindelingen. De vraag luidt dan ook globaal: welke nevenindelingen van de toets acht u zinvol? In het navolgende vindt u **weer een voorstel met de daaraan ten grondslag liggende argumenten**.

Verschillende mogelijkheden zijn de revue gepasseerd om, naast de hoofdingeling volgens de ICPC, als nevenindelingsprincipe voor de kennistoets te hanteren. Overwogen werden indelingen volgens:


1. doodsoorzaken,
2. de inhoud van het cursorisch onderwijs in de beroepsopleiding of de NHG standaarden,
3. leeftijds- en geslachtscategorieën,
4. aspecten van het consult (diagnostiek/beleid),
5. spoedgevallen/niet spoedgevallen,
6. al dan niet fatale consequenties bij foute beantwoording,
7. acute/chronische klachten c.q. aandoeningen, en
8. soorten van kennis (b.v. feitenkennis betreffende nieuwe medicamenten en testen versus interpretatie van klinische gegevens).

Omdat we verwachten dat een groot aantal vragen niet volgens de beide eerste dimensies in te delen zou zijn, vielen deze opties af. Wat betreft de standaarden: dit indelingsprincipe zal pas gehanteerd kunnen worden als over het gehele domein van de huisartsgeneeskunde standaarden beschikbaar zijn. Wel kan per toets worden aangegeven welke vragen betrekking hebben op een uitgebrachte standaard. De derde en vierde indelingsmogelijkheid leken ons inderdaad geschikt om als nevenindeling te hanteren. Met name dient volgens ons gewaakt te worden voor een evenwichtige verdeling per toets tussen vragen over diagnostiek en beleid. Wat betreft de leeftijds- en geslachtsverdeling van patiënten in de toets lijkt het voldoende om over meerdere toetsen heen, bijvoorbeeld gedurende de periode van een jaar, erop te letten dat de verdeling van vragen naar leeftijd en geslacht niet al te scheef komt te liggen. Indelingsprincipe vijf, zes en zeven vielen af, omdat de indeling volgens deze dimensies niet altijd mogelijk is. Niet altijd is bijvoorbeeld uit te maken of een kennisvraag een spoedgeval betreft: de casus kan een spoedgeval betreffen, terwijl de vraag over een (verder wel relevant) aspect gaat dat niet direct betrekking heeft op het spoedeisende karakter van het geval. Zo is ook niet altijd uit te maken of een foute beantwoording (b.v. betreffende een bepaalde handelwijze) de patiënt fataal zou zijn geworden. Tenslotte zijn vele aandoeningen in aanvang acuut maar worden ze bij langer voortduren chronisch. Het laatste indelingsprincipe behelst een interessant aspect, maar vooral vanuit onderzoeksoogpunt en niet zozeer vanuit het gezichtspunt van toetsvragenmaker of -beantwoorder. Het door ons geformuleerde voorstel op basis van de hiervoor beschreven overwegingen luidt om de te hanteren nevenindelingen bij de samenstelling van kennistoetsen te beperken tot **aspecten van het consult** (diagnostiek en beleid) als vaste nevenindeling, en om een evenredige **leeftijds- en geslachtsverdeling per jaar** (over 3 toetsen heen) te bewaken.

VRAAG 3: Bent u het eens met de keuze voor de dimensie 'aspecten van het consult': diagnostiek en beleid als (enige) vaste nevenindeling voor de huisartsgeneeskundige kennistoetsen? Zo nee, wat is uw alternatieve voorstel?

Tot zover onze vragen. Wij hopen met uw medewerking tijdens de groepsdiscussies op basis van de hiervoor beschreven voorstellen tot een goed plan te komen om verantwoorde en acceptabele huisartsgeneeskundige kennistoetsen te kunnen samenstellen.

Bij voorbaat danken wij u heel hartelijk!



Yvonne van Leeuwen
Marjan Potllemans

BIJLAGE OVERZICHT ONDERWERPEN IN DE LANDELIJKE KENNISTOETSEN

juni 88

CARA
KNO bij kinderen
pijn op de borst
hypertensie
diabetes mellitus
buikklachten
oogheelkunde

oktober 88

ritmestoomissen
acute huisartsgen.
radiologie
diversen
besliskunde
mondheelkunde
aand. urinenwegen
KNO

februari 89

CVA
gynaecologie
oogheelkunde
dermatologie
SOA
psychofarmaca
diversen
kinderen

juni 89

lab. diagnostiek
knie en enkel
ulcus pepticum
diabetes mellitus
perifere vaataand.
huisartsenspreekuur
maatsch. factoren
diversen

oktober 89

groei/ontwikkeling
oncologie
alcohol
geriatrie
nek/rugpijn
ritmestoomissen
acute huisartsgen.

februari 90*

oogheelkunde
diabetes mellitus
pijn op de borst
CARA
KNO bij kinderen
hypertensie
buikklachten

juni 90

urologie
atopie/allergie
duizeligh./hoofdpijn
osteoporose
farmacotherapie
obstetrie
bewegingsapparaat
beslisk./epidemiologie

oktober 90

kleine chirurgie
urologie
voeding
psych. problemen
lab. diagnostiek
farm. hart/vaatziekten
oogheelkunde

februari 91

dermatologie
geriatrie
sportgeneeskunde
importziekten
EHBO
vrouw en gezondh.zorg
huisartsenspreekuur

* Eenmaal was het om logistieke redenen ondoenlijk een nieuwe toets samen te stellen. Daarom is de toets van februari 1990 een herhaling van die van juni 1988; zij het in een gewijzigde volgorde van onderwerpen.

BIJLAGE 2 DEFINITIES NEVENINDELINGEN BLAUWDRUK

Definities 'aspecten van het consult'

De definities die werden geformuleerd voor de vier rubrieken van de nevenindeling 'aspecten van het consult' beginnen steeds met een hoofdregel, gevolgd door eventuele bijzonderheden en uitzonderingen.

Diagnostiek

De vraag, c.q. het gevraagde antwoord heeft betekenis in het kader van het diagnostiseren van een aandoening, ziekte of toestandsbeeld.

Hiertoe behoren *wel* vragen/antwoorden betreffende:

- etiologie,
- presentatie van klachten,
- het diagnostiseren van bijwerkingen van geneesmiddelen,
- aanvullende diagnostiek, waarbij de huisarts de aangewezen is om daaruit conclusies te trekken (bijvoorbeeld een door een huisarts aangevraagde echo),
- preventie, indien deze diagnostisch van karakter is.

Hiertoe behoren *niet* vragen/antwoorden die:

- alleen betrekking hebben op frequentie van voorkomen (vallen onder epidemiologie).

Beleid

De vraag, c.q. het gevraagde antwoord heeft betekenis in het kader van het beleid van de aandoening, ziekte of toestandsbeeld. Het betreft zowel het beleid van de patiënt zelf, als van diens omgeving (of adviezen daartoe van de arts), als het beleid van de (huis)arts.

Hiertoe behoren *wel* vragen/antwoorden betreffende:

- het omgaan van de huisarts met klachten,
- 'niets' ondernemen,
- controle afspraken voorstellen/maken,
- preventie, indien deze beleidsmatig van karakter is.

Hiertoe behoren *niet* vragen/antwoorden die:

- uitsluitend gericht zijn op het beloop van de aandoening en/of prognose (vallen onder epidemiologie).

Beleid is onderscheiden in medicamenteus en niet-medicamenteus beleid.

Onder *medicamenteus beleid* wordt verstaan: alle interventies met medicamenten, inclusief suppletie van vitaminen en vaccinaties.

Onder *niet-medicamenteus beleid* wordt verstaan: alle beleid dat niet onder medicamenteus beleid valt, inclusief dieet. Ook adviezen omtrent beleid die indirect op epidemiologische gegevens zijn gebaseerd, vallen onder *deze* categorie.

Epidemiologie

Hiertoe behoren vragen/antwoorden betreffende:

- het vóórkomen van aandoeningen/ziekten/toestandsbeelden,
- het te verwachten beloop (natuurlijk en niet-natuurlijk),
- de prognose,
- risicofactoren.

Hiertoe behoren *wel* vragen/antwoorden betreffende:

- te verwachten complicaties,
- adviezen van de arts waarin expliciet over kansen wordt gesproken.

Hiertoe behoren *niet* vragen/antwoorden:

- die betrekking hebben op etiologie (vallen onder diagnostiek),
- waarbij het voorkomen/het beloop/de prognose een rol speelt, maar die gericht zijn op het stellen van de diagnose of het vaststellen van beleid. In dergelijke gevallen vallen deze vragen onder diagnostiek dan wel beleid. De formulering van de vraag moet hierover uitsluitend geven.

Rest

Vragen/antwoorden die aandoening/ziekte-onafhankelijk zijn (bijvoorbeeld vragen over tuchtrecht).

Definitie 'spoedgeval'

Bij 'spoedgeval' blijft elke gekozen definitie wat kunstmatig. Het gaat eigenlijk om schetsen van gevallen waarbij mag worden aangenomen dat de huisarts zijn of haar bezigheden (spreekuur, lunch etc.) onmiddellijk onderbreekt. Daarom is gekozen voor een omschrijving waarbij vooral uit de tekst van de casus moet blijken of er inderdaad van spoedeisend handelen sprake is.

De volgende definitie werd voor een spoedgeval geformuleerd:

Spoedgeval

De situatie in de geschetste vorm en intensiteit bestaat hoogstens een uur.

Daarbij geldt minstens één van de volgende twee voorwaarden:

- ofwel bestaat er, bij uitblijven van adequaat handelen door wie dan ook, binnen anderhalf uur een reële kans op overlijden of een blijvende handicap, of een toestand/complicatie die intensieve medische bemoeienis vraagt (minstens opname in een ziekenhuis of bemoeienis RIAGG);
- ofwel leidt adequaat ingrijpen binnen anderhalf uur tot opheffing van als ondraaglijk of zeer hevig ervaren angst/pijn.

Tot spoed behoren *niet*:

- situaties waarin angst of pijn er bij de patiënt of bij diens omgeving toe hebben geleid dat met spoed een consult is aangevraagd, terwijl op het moment van aanvragen al duidelijk is dat er medisch gezien geen spoedconsult noodzakelijk is;
- ziekten/aandoeningen waarbij de schade die men oploopt bij uitblijven van adequaat handelen binnen anderhalf uur, doorgaans niet leidt tot blijvende schade.

Checklist 'chronische aandoeningen'

ICPC-code	Aandoening
A t/m Z	alle kwaadaardige ziekten
A t/m Z	congenitale aandoeningen
A t/m Z	chronische allergische aandoeningen
B 81	pemiciëuze anemie
B 83	hemofilie
B 90	aids
D 94	chronische enteritis/colitis ulcerosa
D 97	levercirrhose
F 94	blindheid, alle typen
H 86	doofheid/hardhorendheid
K 74	reuma/reumatoïde hartaandoening
K 75	status na acuut myocardinfarct
K 76	chronische ischaemische hartziekte
K 77	decompensatio cordis
K 83	niet-reumatisch klepgebrek
K 84	cardiomyopathie
K 90	cerebrovasculaire ziekte, excl. TIA
K 91	overige arteriosclerose
L 88	reumatoïde artritis
L 88	M. Bechterew
L 89	artrose
L 95	osteoporose
N	progressieve neurologische aandoeningen
N 86	multiple sclerose
N 87	ziekte van Parkinson
N 88	epilepsie, alle typen
N 89	migraine
P	alle psychosen
P	alle verslavingen
P 70	dementie
R 91	chronische bronchitis/bronchiëctasie
R 95	emfyseem, COPD
R 96	astma
S 87	constitutioneel eczeem
S 91	psoriasis
S 97	ulcus cruris
T	chronische schildklieraandoeningen
T 06	anorexia/bulimia
T 90	diabetes mellitus
U 88	chronische glomerulonefritiden

BIJLAGE 3 GENERALISEERBAARHEIDSTHEORIE

De generaliseerbaarheidstheorie vormt een uitbreiding van de klassieke testtheorie. Volgens deze laatste theorie bestaat een (waargenomen) toetsscore uit twee delen. Een deel wordt bepaald door de *ware score*. Dit is, in tegenstelling tot bij voorbeeld een fysiek kenmerk, een hypothetische, niet observeerbare kwantiteit die alleen indirect kan worden gemeten. Het is de gemiddelde score die iemand zou behalen als hij een oneindig aantal keer opnieuw wordt getoetst, c.q. de score die iemand zou behalen als de meting niet door toevallige omstandigheden wordt beïnvloed (Drenth & Sijsma, 1990). Het overige deel van de waargenomen toetsscore wordt veroorzaakt door (niet systematische) toevalligheden in het meetproces, waarvan wordt aangenomen dat die voor alle deelnemers gelijk zijn, de zogenaamde *random error* (Carmines & Zeller, 1985). Naarmate de error groter is, ligt de ware score verder van de waargenomen score af, en is de score dus minder betrouwbaar.

De klassieke testtheorie maakt gebruik van het feit dat een toets door meerdere deelnemers wordt gemaakt, zodat de spreiding van de scores kan worden gebruikt bij de betrouwbaarheidsbepaling. Als maat voor de spreiding wordt de *variantie* gebruikt. De variantie van de waargenomen scores is gelijk aan de variantie van de ware scores plus de variantie van de meetfout (dit is de toevals- of errorvariantie). De betrouwbaarheid wordt gedefinieerd als de verhouding tussen de variantie van de ware scores (ofwel 'ware variantie') en de variantie van de waargenomen scores (ofwel 'waargenomen variantie'). Dit levert een correlatiecoëfficiënt op, de zogenaamde *betrouwbaarheidscoëfficiënt*. Deze coëfficiënt varieert tussen 0 (geen correlatie tussen ware en waargenomen score) en 1 (maximale correlatie tussen ware en waargenomen score), en neemt een lagere waarde aan naarmate de errorvariantie groter is (Thorndike, 1971). Omdat de ware score van de deelnemers aan een toets alleen indirect kan worden bepaald, kan de betrouwbaarheid alleen bij benadering worden vastgesteld. De schatting van de betrouwbaarheid van toetsscores berust op het concept van parallelle metingen, waarbij wordt gezocht naar een maat voor de reproduceerbaarheid van de toetsscores. Idealiter wordt de betrouwbaarheidscoëfficiënt bepaald door de variantie te bepalen van de scores op twee toetsafnames onder identieke omstandigheden. Dit is echter in de praktijk onmogelijk. Daarom zijn er verschillende schattingsmethodes ontwikkeld die elk leiden tot een bepaalde betrouwbaarheidscoëfficiënt (Meerling, 1981). Zo kan de betrouwbaarheid worden geschat met behulp van methoden die zijn gebaseerd op herhaalde toetsing, en met behulp van methoden waarin bewerkingen plaatsvinden op basis van de scores behaald op een enkele toets. Een voorbeeld van een betrouwbaarheidsberekeningsmethode waarbij slechts één meting wordt gebruikt, is de methode die is gebaseerd op itemanalyse. Deze methode leidt tot een zogenaamde coëfficiënt van itemconsistentie (ofwel homogeniteitscoëfficiënt, ook wel bekend als Cronbach's alpha) (Drenth 1975).

Welke methode de meest geëigende is om de betrouwbaarheidscoëfficiënt te schatten, is afhankelijk van de aard van de test en het te toetsen kenmerk. In de klassieke testtheorie worden, buiten de variantiebron die leidt tot 'ware variantie', dat wil zeggen de verschillen tussen individuen, geen andere variantiebronnen onderscheiden. Alle overige factoren die tot verschillen in toetsscores kunnen leiden, worden beschouwd als bijdragen aan de *errorvariantie*. Deze kan niet nader worden geïdentificeerd (Crocker & Algina, 1986).

De *generaliseerbaarheidstheorie* gebruikt hetzelfde theoretische kader als de klassieke testtheorie, maar vanuit een breder perspectief. Daarbij gaat men ervan uit, dat voor verschillende gebruiksdoelen van een toets het *object van meting*, dat wil zeggen de variantiebron die leidt tot 'ware' variantie, kan variëren. In plaats van de term 'ware' variantie wordt dan ook de term '*bedoelde*' variantie gehanteerd. Bedoelde variantie is dus dat deel van de variantie waar de aandacht van de onderzoeker op is gericht, namelijk de variantie veroorzaakt door het object van meting. Dit kan de kennis van individuen zijn, maar ook het gemiddelde kennisniveau van bepaalde groepen van individuen (Suen, 1990; van der Vleuten & Wijnen, 1991). Wat wordt beschouwd als *errorvariantie* (dit is de niet bedoelde variantie) is eveneens afhankelijk van het gebruiksdoel van de toets. Analyses kunnen alleen worden uitgevoerd als bekend is welke variantiebron wordt beschouwd als bedoelde variantie en welke bronnen worden

beschouwd als errorvariantie. Daarmee geeft de generaliseerbaarheidstheorie inzicht in de mate waarin verschillende bronnen bijdragen aan de toetsvariantie. De beoogde gebruiksfunctie voor de toets bepaalt volgens de generaliseerbaarheidstheorie zowel het object van meting als datgene wat moet worden beschouwd als errorvariantie. Als een toets verschillende gebruiksfuncties heeft, kan zowel het object van meting, en daarmee de bron van bedoelde variantie, als de errorvarianties variëren. Er kunnen vervolgens verschillende betrouwbaarheden worden bepaald voor verschillende gebruiksfuncties van dezelfde toets (Suen, 1990; van der Vleuten & Wijnen, 1991). Alle potentiële variantiebronnen, behalve de variantiebron behorend bij het object van meting, worden in de generaliseerbaarheidstheorie aangeduid met de term '*facet*'. Als er in een generaliseerbaarheids-onderzoeksdesign, naast het object van meting, nog één andere variantiebron is opgenomen, is er sprake van een '*one-facet*' design. Zijn er in het onderzoeksdesign twee andere bronnen opgenomen, dan is er sprake van een '*two-facet*' design (Suen, 1990). Facetten kunnen *random* of *fixed* zijn. Een '*random facet*' bestaat uit elementen die worden beschouwd als een random steekproef van alle mogelijke elementen die deel zouden kunnen uitmaken van dat facet. Bijvoorbeeld: als het facet items wordt beschouwd als random, betekent dit dat bij een nieuwe toetsafname een andere steekproef van items wordt getrokken. Een facet wordt als *fixed* beschouwd, als bij alle toekomstige meetprocedures gebruik wordt gemaakt van exact dezelfde elementen van het betreffende facet (Suen, 1990). Een '*fixed facet*' perkt het object van meting in. Als bijvoorbeeld een groep items wordt beschouwd als *fixed*, betekent dit dat deze items geen steekproef vormen uit een bepaald domein, maar dat ze zelf met elkaar het gehele domein vormen. De groep van items waaruit de toets bestaat, had er niet 'toevallig' anders kunnen uitzien. De variantie die wordt veroorzaakt door de interactie met het object van meting en het '*fixed facet*' wordt daarmee onderdeel van het object van meting. Deze interactie kan, omdat de items worden beschouwd als onveranderlijk gegeven, immers niet meer tot de errorvariantie worden gerekend. Als een facet als '*fixed*' wordt beschouwd, zal de betrouwbaarheid dan ook hoger uitvallen dan wanneer er sprake is van een *random facet*. Naarmate meer facetten '*fixed*' zijn, kunnen er uit de betrouwbaarheidsstudie minder vergaande conclusies worden getrokken. Ter toelichting: als in het vorige voorbeeld verschillen in score tussen personen het object van meting vormen, heeft de gevonden betrouwbaarheid van de scores slechts betrekking op de gevonden verschillen in score voor wat betreft de groep van items waarop de berekeningen zijn gebaseerd. Om de generaliseerbaarheidstheorie in een betrouwbaarheidsstudie te kunnen toepassen, moet er minstens één facet *random* zijn (Suen, 1990).

Een betrouwbaarheidsstudie, die wordt uitgevoerd op basis van de generaliseerbaarheidstheorie, bestaat uit twee onderdelen: een generaliseerbaarheidsstudie (*G-studie*) en een of meer besluitvormingsstudies (*decision study: D-studie*). De basis van een *G-studie* is een variantie-analyse. In de *G-studie* worden vervolgens zo veel mogelijk variantiebronnen, dat wil zeggen zowel bronnen die leiden tot bedoelde als bronnen die leiden tot errorvariantie, in kaart gebracht. Van deze variantiebronnen wordt de relatieve bijdrage aan de totale variantie geschat en uitgedrukt in variantiecomponenten (Brennan, 1983). In *D-studies* kunnen, op basis van de variantiecomponenten uit de *G-studie*, diverse betrouwbaarheidsschattingen worden uitgevoerd. In de *D-studies* kunnen zo de consequenties voor de betrouwbaarheid worden geschat van veranderingen in het meetdesign. Zo kan er worden gevarieerd met het aantal elementen dat deel uitmaakt van een bepaald facet (zoals aantallen items) of met groeps grootte, om vervolgens de invloed op de betrouwbaarheid van eventuele wijzigingen te onderzoeken. Ook kan de invloed op de betrouwbaarheid worden onderzocht, als een bepaald facet in plaats van '*random*' wordt beschouwd als '*fixed*'. Een andere toepassingsmogelijkheid van een *D-studie* is om na te gaan wat de te bereiken betrouwbaarheid is vanuit verschillende perspectieven op het gebruik van de scores, dus voor het normgeoriënteerde (relatieve), het domeingeoriënteerde (absolute) en het mastery georiënteerde (beslissings) perspectief. Door op deze manier met het aantal items, c.q. het aantal personen, te variëren in het meetdesign, kunnen schattingen worden gemaakt voor het noodzakelijke aantal items dat een toets moet hebben om, vanuit de verschillende gebruiksperspectieven, een bepaalde betrouwbaarheid te bereiken (Eggen & Sanders, 1993). Daarom wordt de generaliseerbaarheidstheorie ook wel een theorie genoemd, die specifiek op praktisch gebruik is afgestemd (van der Vleuten & Wijnen, 1991).

BIJLAGE 4 FORMULES TOEGEPAST IN BETROUWBAARHEIDSSCHATTINGEN

normgeoriënteerd (relatief) betrouwbaarheidsperspectief:

variantie tgv object van meting

g-coëfficiënt:

totale variantie (som van alle variantiecomponenten) behalve de variantie ten gevolge van de moeilijkheidsgraad van de items

$$G = \frac{p}{p + \frac{pi}{n_i}}$$

domeingeoriënteerd (absoluut) betrouwbaarheidsperspectief:

variantie tgv object van meting

phi-coëfficiënt:

totale variantie (som van alle variantiecomponenten)

$$\Phi = \frac{p}{p + \frac{i}{n_i} + \frac{pi}{n_i}}$$

beslissingsgeoriënteerd betrouwbaarheidsperspectief:

variantie tgv object van meting plus kwadraat (gemiddelde score - grenswaarde)

adjusted phi:

totale variantie plus kwadraat (gemiddelde score - grenswaarde)

$$\Phi_{\lambda} = \frac{p + (\mu - \lambda)^2}{p + \frac{i}{n_i} + \frac{pi}{n_i} + (\mu - \lambda)^2}$$

SEM: wortel uit de errorvariantie

$$SEM = \sqrt{\frac{pi}{n_i}}$$

Formule SEM voor groepsgemiddelden (relatief perspectief):

$$SEM = \sqrt{\frac{p}{n_p} + \frac{pi}{n_p \times n_i}}$$

Formule SEM voor groepsgemiddelden (absoluut perspectief):

$$SEM = \sqrt{\frac{p}{n_p} + \frac{i}{n_i} + \frac{pi}{n_p \times n_i}}$$

**BIJLAGE 5 VOORBEELDPAGINA ALGEMENE HUISARTSGENEESKUNDIGE
KENNISTOETS FEBRUARI 1992**

Landelijke Experimentele Kennistoets februari 1992

Oog (ICPC F)

Mevrouw Lafrijsen, 45 jaar, komt bij de huisarts omdat ze sedert ongeveer een uur witte lichtflitsen voor het linker oog ziet, die constant aanwezig zijn en verergeren wanneer ze naar rechts kijkt. Ze heeft geen hoofdpijn en ze is niet bekend met migraine of epilepsie. Ze is brildragend. Visusonderzoek geeft de volgende uitslag (met correctie): VOS 1.0, VOD 1.0. Het onderzoek van het voorste oogsegment is zonder afwijkingen. Bij funduscopie krijgt de huisarts alleen de papil in beeld; deze is niet afwijkend.

Bovenstaande bevindingen passen bij de diagnose(n):

- 131. glasvochtmembraanloslating; [juist]
- 132. mouches volantes (eiwitcondensaten in het glasvocht); [onjuist]
- 133. netvliesscheur. [juist]

Lit. vr. 131-133: Hollwich F. Leerboek oogheelkunde 1989: 166.

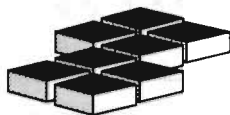
Meneer Jochems, 38 jaar, komt op het spreekuur omdat zijn wimpers elke ochtend aan elkaar geplakt zitten. Hij heeft deze klacht sedert 3 weken. Hij heeft al chlooramfenicol-oogzalf gebruikt, die zijn vrouw nog had liggen, echter zonder resultaat. De huisarts constateert een blepharitis. Zij vervangt de chlooramfenicol-oogzalf door neomycine-oogdruppels.

- 134. Dit is in dit geval een correcte therapie. [onjuist]

Meneer Jochems vraagt hoe hij nou aan een dergelijke "ooglidrandonsteking" gekomen kan zijn. De huisarts zegt dat één van de oorzaken van een blepharitis een slecht gecorrigeerde refractie-afwijking is.

- 135. Dit is correcte informatie. [juist]

Lit. vr. 134 en 135: Henkes HE, Van Balen AThM. Oogheelkunde. Amsterdam/Brussel: Elsevier 1983: 94.

BIJLAGE 6: VRAGENLIJST ACHTERGRONDKENMERKEN**Samenwerkingsverband Interfacultair Overleg Huisartsgeneeskunde**

Uitvoerend Bureau

deelnemerscode:

--	--	--	--	--	--	--	--

VRAGENLIJST HUISARTSEN FEBRUARI 1992

gelieve deze
rubriek invullen³

PERSOONLIJKE KENMERKEN

1. Geslacht: man ☐ vrouw ☐
2. Geboortedatum:-.....-19....

PROFESSIONELE KENMERKEN**Opleidingskenmerken**

3. Heeft u de beroepsopleiding tot huisarts gevolgd: ja ☐
nee ☐
(indien nee: ga door met vraag 6)
4. Waar heeft u de beroepsopleiding tot huisarts gevolgd:
- a. Amsterdam (UvA) ☐
 - b. Amsterdam (VU) ☐
 - c. Groningen ☐
 - d. Leiden ☐
 - e. Maastricht ☐
 - f. Nijmegen ☐
 - g. Rotterdam ☐
 - h. Utrecht ☐
 - i. elders, namelijk:.....

³ antwoorden graag aankruisen of invullen.

NB: bij sommige vragen kunnen meerdere antwoorden van toepassing zijn.

5. Duur beroepsopleiding:

1 jaar ☐

2 jaar ☐

Werkzaamheden als praktiserend huisarts

6. Hoeveel jaar bent u werkzaam als huisarts:

aantal jaar:

(bedoeld worden werkzaamheden in eigen praktijk, in dienstverband in een huisarts-praktijk, vaste waarneming)

7. Hoeveel dagen per week bent u gemiddeld werkzaam als huisarts (excl. weekends):

gemiddeld aantal dagen:

8. Van welke beroepsvereniging bent u lid:

a. NHG ☐

b. LHV ☐

c. geen ☐

d. anders, namelijk

Onderwijservaring

9. Heeft u ervaring als opleider/docent op het gebied van de huisartsgeneeskunde:

a. in de basisopleiding ☐

b. in de beroepsopleiding ☐

c. in de nascholing ☐

d. nee ☐

(indien nee: ga door met vraag 11)

10. Totaal aantal jaar onderwijservaring:

....

Nascholingskenmerken

11. Hoeveel **dagdelen** (middag, ochtend, avond) heeft u in het **afgelopen jaar (1991)** besteed aan buiten uw praktijk georganiseerde **nascholing**:

aantal dagdelen:

12. Hoeveel uur besteedt u gemiddeld per week aan het **lezen van vakliteratuur** (boeken en tijdschriften):

aantal uur per week:

13. Bent u lid van een vaste **toetsingsgroep** van huisartsen (met regelmatige bijeenkomsten gericht op toetsing van eigen handelen):

ja ☐

nee ☐

PRAKTIJKKENMERKEN

14. In wat voor soort van **werkverband** oefent u uw praktijk uit:
- | | |
|---------------------------|--------------------------|
| a. solopraktijk | <input type="checkbox"/> |
| b. duopraktijk/associatie | <input type="checkbox"/> |
| c. groepspraktijk | <input type="checkbox"/> |
| d. gezondheidscentrum | <input type="checkbox"/> |
| e. anders, namelijk | |
15. Waar is uw praktijk **geografisch gesitueerd**:
- | | |
|---|--------------------------|
| a. grote stad (> 100.000 inwoners) | <input type="checkbox"/> |
| b. stad (tussen 30.000 en 100.000 inwoners) | <input type="checkbox"/> |
| c. verstedelijkt platteland (tussen 5.000 en 30.000 inwoners) | <input type="checkbox"/> |
| d. platteland (< 5.000 inwoners in grootste woonkern) | <input type="checkbox"/> |
16. **Hoeveel patiënten** (ziekenfonds plus particulier verzekerden) telt uw **praktijk**:
- aantal patiënten:
17. **Hoeveel patiëntcontacten heeft u gemiddeld per week** (inclusief diensten):
- aantal patiëntcontacten per week:
18. Heeft uw **patiëntenpopulatie** bepaalde kenmerken waardoor zij zich onderscheidt van een doorsnee-praktijkpopulatie?
- | | |
|------------------------------------|--------------------------|
| a. nee | <input type="checkbox"/> |
| b. relatief veel ouderen | <input type="checkbox"/> |
| c. relatief veel jonge gezinnen | <input type="checkbox"/> |
| d. anders (graag toelichten):..... | |
| | |

EXPERTISEGEBIEDEN

19. Elke huisarts heeft zo zijn of haar eigen speciale aandachtsgebieden, belangstellingen of stokpaardjes. De een kan bijvoorbeeld veel weten van diabetes, de ander gebruikt speciale apparatuur, zoals ECG's of oogheelkundige apparatuur. Wij vragen u hieronder om per onderwerp aan te geven op welke gebieden uw expertise naar uw eigen idee groter, ongeveer gelijk dan wel kleiner is dan die van uw collega's.

onderwerpen **relatieve expertise**

Klachten en aandoeningen op het gebied van:

	veel	ge- middeld	weinig
a. het bloed en bloedvormende organen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. de tractus digestivus	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. het oog	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. het oor	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. de tractus circulatorius	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. het bewegingsapparaat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g. het zenuwstelsel	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h. psychosociale klachten en aandoeningen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i. de tractus respiratorius	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
j. de huid	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
k. endocriene klieren/metabolisme/voeding	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
l. de urinewegen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
m. zwangerschap/bevalling/anticonceptie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
n. de vrouwelijke geslachtsorganen (incl. mammae)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
o. de mannelijke geslachtsorganen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

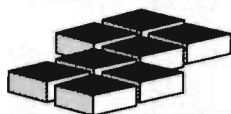
Tot zover de vragenlijst.

Wij danken u van harte voor uw bereidwilligheid de vragen te beantwoorden.

BIJLAGE 7 SATISFACTIEVRAGENLIJST KENNISTOETS FEBRUARI 1992

Samenwerkingsverband Interfacultair Overleg Huisartsgeneeskunde

Uitvoerend Bureau



SATISFACTIE VRAGENLIJST LANDELIJKE KENNISTOETS

februari 1992, toetsnummer: 9202

Dit formulier dient met het antwoordformulier van de kennistoets te worden ingeleverd.

Deelnemerscode:

--	--	--	--	--	--	--	--

omcirkelen

	geheel eens	eens	neutraal	oneens	geheel oneens	niet te beoordelen/ geen mening
1. Ik vond het prettig om deze kennistoets te maken:	1	2	3	4	5	6
2. Ik beoordeel de onderwerpen, waaruit de toets bestond, over het geheel genomen als huisartsgeneeskundig relevant:	1	2	3	4	5	6
3. Ik vond de toets te lang:	1	2	3	4	5	6
4. Ik beoordeel de vragen, waaruit de toets bestond, over het geheel genomen als huisartsgeneeskundig relevant:	1	2	3	4	5	6
5. Ik vond het lastig om 'Juist-Onjuist-Vraagteken'-vragen te beantwoorden:	1	2	3	4	5	6
6. Ik vind deze toets een bruikbaar hulpmiddel om mijn eigen kennis-lacunes vast te stellen:	1	2	3	4	5	6

7. Ik vind deze toets goed bruikbaar als uitgangspunt voor het bepalen van mijn eigen onderwijs- of (na)-scholingsprogramma:	1	2	3	4	5	6
	<½u.	½-1u.	1-1½u.	1½-2u.	>2u.	
8. Hoeveel tijd heeft u nodig gehad om deze kennistoets te maken?	1	2	3	4	5	

OVERIG COMMENTAAR OP DE KENNISTOETS:

1. Ten aanzien van de toets als geheel:

2. Ten aanzien van de toetsvragen:
(s.v.p. vraagnummers vermelden)

BIJLAGE 8 INSTRUCTIEBRIEF HUISARTSEN-DEELNEMERS LANDELIJKE KENNISTOETS FEBRUARI 1992

Aan: huisartsen die meedoen aan de landelijke kennistoets van februari 1992

15 januari 1992

US92004.MP

Landelijke kennistoets februari 1992

Geachte heer/mevrouw,

Met deze brief beogen wij u te voorzien van de voor u noodzakelijke informatie over de landelijke huisartsgeneeskundige kennistoets. Deze toets wordt aan u voorgelegd in het kader van een onderzoek dat eind 1990 is ingesteld naar onder meer factoren die het kennisniveau van huisartsen en huisartsen in opleiding bepalen. Uitgebreidere achtergrond-informatie over de huisartsgeneeskundige kennistoets treft u aan in bijgesloten brochure 'De kennistoets in de beroepsopleiding tot huisarts'.

HET KENNISONDERZOEK

Het onderzoek kent twee onderdelen:

- één deelonderzoek is gericht op de kennisontwikkeling in de beroepsopleiding en op een aantal verschillende factoren die daarop, naar wordt verwacht, invloed uitoefenen. Het betreft factoren als het aangeboden instituutsonderwijs, de huisartsopleider, zelfstudie, kenmerken van de opleidingspraktijk etc.
- het andere deelonderzoek is erop gericht te achterhalen wat de bijdrage van de kennistoets kan zijn in de deskundigheidsbevordering en nascholing van gevestigde huisartsen.

Het onderzoek wordt onder auspiciën van NHG en LHV uitgevoerd door medewerkers van de Werkgroep Onderzoek Kwaliteit Huisartsgeneeskunde (Rijksuniversiteit Limburg Maastricht) en door medewerkers van het landelijk Samenwerkingsverband-IOH (Rijksuniversiteit Utrecht).

Aan de kennistoets van februari 1992 nemen verschillende groeperingen deel: co-assistenten huisartsgeneeskunde, huisartsen in opleiding, en gevestigde huisartsen, al dan niet als opleider bij de beroepsopleiding betrokken.

Het vervolg van deze brief heeft betrekking op enkele praktische zaken die u moet weten om de toetsvragen te kunnen beantwoorden.

PRAKTISCHE INFORMATIE KENNISTOETS FEBRUARI 1992

- de toetsbescheiden

In uw 'toetspakket' treft u de volgende bescheiden aan:

- een rood toetsboekje;
- een antwoordformulier met doorslagvel;
- een vragenlijst naar uw achtergrondkenmerken;
- een satisfactieformulier;
- een retourenveloppe; en ten slotte
- de brochure over de achtergrond van de kennistoets.

De verschillende formulieren zijn voorgecodeerd teneinde verwerking van de gegevens mogelijk te maken.

- toetsvragen en antwoordformulier (zie ook pag. 3 t/m 5 van het rode toetsboekje)

De kennistoets bestaat uit 160 toetsvragen, meestal in de vorm van een casus waarbij één of meer **stellingen** behoren. Deze stellingen vormen de eigenlijke toetsvragen, die moeten worden beantwoord. Daarvoor bestaat per stelling de keuze uit de antwoordcategorieën 'juist', 'onjuist' of 'vraagteken'. Door een stelling met een vraagteken te beantwoorden geeft u aan het antwoord niet te weten. (N.B.: het vraagteken dient dus niet om aan te geven hoe zeker u van uw antwoord bent!).

De antwoorden moeten **per stelling MET POTLOOD** op het bijgevoegde **antwoordformulier** worden aangestreept. Het antwoordformulier is een doordrukformulier in duplo, waarvan u de onderste bladzijde zelf kunt houden.

- anonimiteit

De lijst met namen van deelnemers aan de toets en hun bijbehorende codenummers is alleen bekend op het Uitvoerend Bureau van het landelijk Samenwerkingsverband van het IOH waar de verwerking van de antwoordformulieren en vragenlijsten zal plaatsvinden. Uw toetsresultaten krijgt u ook via ditzelfde Bureau terug.

- vragenlijsten

Behalve toetsboekje en antwoordformulier treft u nog twee vragenlijsten aan.

De 'Vragenlijst huisartsen februari 1992' is bedoeld om nadere informatie te verzamelen over enkele kenmerken van uzelf en uw praktijk die mogelijk van belang zijn bij de interpretatie van de toetsresultaten op groepsniveau. Wij verzoeken u daarom deze vragenlijst ingevuld met het antwoordformulier te retourneren.

De tweede lijst is een korte 'Satisfactie-vragenlijst' waarop u via het aankruisen van enkele vijfpuntsstellingen onder andere uw oordeel kunt geven over de mate van huisartsgeneeskundige relevantie van de toetsonderwerpen en -vragen. Op de achterkant van dit formulier kunt u bovendien opmerkingen kwijt over toetsvragen die onverhoopt onduidelijk of dubbelzinnig gesteld waren, over de relevantie van bepaalde vragen etc. Ook deze lijst vragen wij u zorgvuldig in te vullen en met de overige formulieren te retourneren.

- procedures

Omdat een aantal deelnemende huisartsen de kennistoets thuis invult en andere deelnemers de toets op het instituut, onder soortgelijke condities als de huisartsen in opleiding, invullen, bestaan er twee verschillende procedures met betrekking tot de afwikkeling van de formulieren.

Voor de deelnemers die de toets **thuis** invullen geldt de volgende procedure:

Wij verzoeken u om, nadat u het antwoordformulier en de vragenlijsten hebt ingevuld, deze door middel van de bijgevoegde, reeds geadresseerde retourenveloppe, toe te sturen aan het Samenwerkingsverband te Utrecht. Wellicht ten overvloede willen we u er hierbij toch nog op wijzen dat het van **wezenlijk belang** is dat u de kennistoets zoveel mogelijk onder '**examencondities**' invult: dwz **zonder naslagwerken en zonder overleg**. Alleen dan zijn de toetsresultaten interpreteerbaar.

De ingevulde formulieren moeten **uiterlijk 28 februari** in Utrecht zijn. Wij vragen u dan ook om de formulieren **uiterlijk 26 februari a.s.** te verzenden. U kunt de antwoordsleutel van de toets verwachten in de week van 28 februari.

Voor de deelnemers die de toets **op het instituut** invullen geldt de volgende procedure:

U ontvangt alle toetsbescheiden via het stafid dat op uw instituut de afname van de kennistoets coördineert. De ingevulde antwoordformulieren en vragenlijsten kunt u bij ditzelfde stafid weer inleveren. Hij of zij zal vervolgens ervoor zorgen dat de bescheiden tijdig in Utrecht arriveren voor de verwerking. Als het moment van toetsafname interfereert met de toetsafname bij de huisartsen in opleiding, zal op het instituut dezelfde procedure worden gehanteerd als bij de huisartsen in opleiding. **Dat wil zeggen dat de toetsboekjes na de toetsafname tijdelijk worden ingenomen tot het moment waarop de antwoordsleutel wordt verspreid, dwz tot eind februari. Vanaf dat moment kunt u via het betrokken stafid uw eigen toetsboekje plus antwoordsleutel weer terugkrijgen.**

- terugrapportage

Ongeveer zes weken nadat u de kennistoets hebt gemaakt, kunt u een overzicht verwachten van uw eigen, individuele toetsresultaat, naast de gemiddelde resultaten van alle huisartsen die aan de kennistoets deelnemen. Als u de toets invult als lid van een bepaalde groep (bv een toetsingsgroep, of een groep huisartsopleiders), dan krijgt u ook een overzicht van de gemiddelde resultaten van deze groep.

Behalve een totaal-overzicht ontvangt u de resultaten ook gespecificeerd per onderwerp uit de toets.

Wij danken u bij voorbaat heel hartelijk voor uw bereidheid aan de landelijke huisartsgeneeskundige kennistoets van februari 1992 te willen deelnemen.

We hopen dat u veel plezier beleeft aan de beantwoording van de toetsvragen. Mocht u naar aanleiding van deze brief of naar aanleiding van de toets nog vragen of opmerkingen hebben, dan zijn wij natuurlijk altijd bereid daarop in te gaan.

Met vriendelijke groeten,

(mede namens alle overige betrokkenen bij de ontwikkeling en het onderzoek van de huisartsgeneeskundige kennistoets)



MC Poffamans, arts/onderwijskundige

YB van Leeuwen, huisarts

SSL Mol, huisarts

PS:

Wellicht is er in uw praktijk een co-assistent of een huisarts in opleiding aanwezig, die de landelijke kennistoets ook maakt (op het instituut). In dat geval verzoeken wij u om geheimhouding van de inhoud van de toets, tot het moment waarop de co-assistent, c.q. huisarts in opleiding de toets zelf gemaakt heeft.

Bijlagen:

- toetsboekje: landelijke experimentele kennistoets februari 1992
- voorgecodeerd antwoordformulier
- voorgecodeerde vragenlijst achtergrondkenmerken
- satisfactievragenlijst
- brochure: 'De kennistoets in de beroepsopleiding tot huisarts'
- retourenveloppe

BIJLAGE 9 VOORBEELD INDIVIDUELE FEEDBACK KENNISTOETS FEBRUARI 1992, PLUS TOELICHTENDE BRIEF

Aan: de huisartsen die hebben deelgenomen aan de landelijke kennistoets van februari 1992

29 maart 1992

US92116.MP

Feedback kennistoets februari 1992

Geachte heer, mevrouw,

Hierbij ontvangt u uw persoonlijke feedback op de landelijke huisartsgeneeskundige kennistoets waaraan u in februari j.l. heeft deelgenomen. De antwoordsleutel (de lijst met de antwoorden op de toetsvragen) ontving u al enkele weken geleden.

De feedback bestaat uit twee formulieren waarop uw individuele toetsresultaten zijn weergegeven alsmede enkele resultaten op groepsniveau. De inhoud van beide formulieren lichten we hierbij toe.

Links bovenaan ziet u op beide formulieren de datum (februari 1992) en de codenaam van de toets (LEK9202). Daaronder ziet u uw persoonlijke code (kandidaatnummer); het aantal 'geldige items' (146; voor toelichting zie hierna); de omvang van de groep waartoe u behoort (b.v. opleidersgroep, toetsgroep); en tenslotte het totale aantal huisartsen dat aan de toets van februari 1992 heeft deelgenomen (365).

- de resultaten per toetsonderwerp

Op het eerste formulier treft u de 17 onderwerpen aan waaruit de kennistoets bestond. Naast de kolom waarin deze onderwerpen zijn genoemd is het aantal vragen per onderwerp weergegeven. In de kolom daarnaast treft u per onderwerp uw eigen individuele toetsresultaat. De resultaten zijn weergegeven in het percentage van de vragen dat u per onderwerp goed, fout of met een vraagteken heeft beantwoord. Ook is weergegeven een procentuele score (G-F) waarin de foute antwoorden van de goede zijn afgetrokken.

Naast de kolom waarin uw persoonlijke toetsresultaat is weergegeven, kunt u de gemiddelde scores zien van uw groepsgenoten; in de laatste, meest rechtse kolom treft u de gemiddelde resultaten aan van alle huisartsen die aan de toets hebben deelgenomen.

Uiterst rechts is weergegeven of uw persoonlijk resultaat op het betreffende onderwerp relatief 'hoog' of relatief 'laag' was ten opzichte van de gemiddelden van de totale groep deelnemende huisartsen. Voor de inschatting 'hoog' resp. 'laag' is als uitgangspunt genomen het percentage goed beantwoorde vragen minus het percentage fout beantwoorde vragen. Als grens voor 'hoog', resp. 'laag' is de standaarddeviatie genomen: als uw goed-minus-foutscore meer dan één standaarddeviatie boven het groepsgemiddelde uitkomt, is uw score 'hoog' genoemd. Als deze score meer dan één standaarddeviatie beneden het groepsgemiddelde is uitgevallen, is dat als 'laag' bestempeld. (NB: uit de hiervoor

beschreven toelichting zal u duidelijk zijn dat 'hoog' of 'laag' geen absolute criteria weergeven. Uiteraard speelt daarnaast bij de interpretatie van de gegevens ook het aantal vragen dat over het betreffende onderwerp in de toets terecht is gekomen, een rol.)

Onder de onderwerpenlijst op dit eerste formulier vindt u uw persoonlijke totaalresultaat met daarnaast de gemiddelde totaalresultaten op groepsniveau.

Onderaan het formulier zijn de nummers aangegeven van de vragen die niet zijn meegenomen bij de berekening van de toetsresultaten. Uit de toets van februari 1992, die oorspronkelijk uit 160 vragen bestond, zijn namelijk 14 vragen verwijderd op grond van commentaren van deelnemers aan de toets en/of op grond van de statistische gegevens die per vraag zijn berekend.

- de resultaten per nevenindeling

Op het tweede formulier treft u nogmaals een weergave van uw persoonlijke toetsresultaten en de gemiddelden van uw groepsgenoten, maar nu weergegeven vanuit een andere invalshoek.

Dwars door de toets (dwz los van de onderwerpenindeling) is er bij de samenstelling van de toets voor gezorgd dat deze een redelijk aandeel vragen betreffende diagnostiek, beleid en epidemiologische aspecten bevatte. Daarnaast is ervoor gezorgd dat in de toetscasuïstiek voldoende spreiding naar voren kwam over verschillende leeftijdscategorieën van patiënten. De feedback is daarom nog eens ingedeeld naar de categorieën die bij deze alternatieve toetsindeling zijn gehanteerd.

Tot zover onze toelichting bij uw feedback. Mocht u naar aanleiding van de toets of de resultaten nog vragen hebben, dan kunt u zich daarmee altijd tot ons richten.

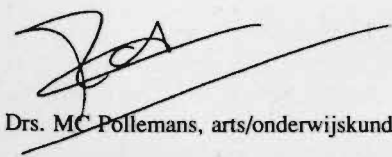
Wij danken u van harte voor uw bereidwilligheid de toetsvragen te beantwoorden. Wij hopen dat u de kennistoets met plezier heeft gemaakt. Als 'tegenprestatie' voor de door u geleverde moeite voor het invullen van de toetsvragen en de bijbehorende vragenlijsten, kunnen wij u een bedrag van f. 50,- aanbieden. Om dit bedrag aan u te kunnen overmaken, hebben wij echter enkele nadere gegevens van u nodig. Als u deze invult op bijgaand antwoordstrookje en dit strookje in de bijgesloten retourenveloppe terugstuurt, volgt uitbetaling via de Rijksuniversiteit Limburg.

Mocht u, c.q. de toetsings- of opleidersgroep waartoe u behoort, naar aanleiding van uw deelname aan de kennistoets van februari 1992 belangstelling hebben voor frequentere deelname aan landelijke kennistoetsen, dan is dat in principe mogelijk. Wij verzoeken u daartoe contact op te nemen met het Uitvoerend Bureau van het landelijk Samenwerkingsverband te Utrecht (gevestigd bij de Vakgroep Huisartsgeneeskunde, Bijlhouwerstraat 6, 3511 ZC Utrecht; tel.nr. 030-331123).

Ook als u belangstelling heeft voor kennistoetsing op meer specifieke (onderwerps)gebieden, kunt u daar voor nadere informatie terecht.

Wij hopen u hiermee van voldoende informatie te hebben voorzien. Namens de onderzoekers en overige betrokkenen bij de ontwikkeling van de landelijke kennistoets,

hoogachtend,



Drs. MC Pollemans, arts/onderwijskundige

HUISARTSGENEESKUNDE, SV-IOH, UTRECHT
KENNISTOETS HUISARTSEN

Toetsdatum : FEBRUARI 1992
Toetsnaam : Kennistoets feb. 1992
Kandidaat nummer : Naam: Jaar:
Aantal geldige items : 146
Eigen groep : Huisartsen (niet-opleiders) (n=133)
Deeln. landel. : Alle landelijke deelnemende huisartsen (n=365)
UW RESULTAAT IS :

Toetsresultaten per categorie. Per individu en voor twee groepen (percentages)

CATEGORIE OMSCHRIJVING	AANT. VRAGEN	INDIVIDUEEL				EIGEN GROEP, n=133				DEELN. LANDEL., n=365				-/o/+
		GOED	FOUT	?	G-F	GOED	FOUT	?	G-F	GOED	FOUT	?	G-F	
ALGEMEEN EN NIET GESPECIF	7	85.7	0.0	14.3	85.7	69.1	19.2	26.0	49.8	67.6	20.3	26.3	47.3	+
BLOED EN BLOEDVORMENDE OR	2	100.0	0.0	0.0	100.0	82.7	14.7	2.6	68.0	81.8	15.8	2.5	66.0	+
TR.DIGESTIVUS	9	77.8	22.2	0.0	55.6	71.3	20.7	19.1	50.5	70.7	22.9	17.5	47.9	0
OOG	6	66.7	33.3	0.0	33.3	59.1	33.2	7.6	25.9	60.0	34.4	5.6	25.6	0
OOR	5	60.0	0.0	40.0	60.0	68.9	18.3	32.8	50.5	69.9	19.2	31.0	50.7	0
TR.CIRCULATORIUS	14	64.3	21.4	14.3	42.9	69.4	21.6	23.3	47.9	71.5	20.5	22.2	51.0	0
BEWEGINGSAPPARAAT	18	50.0	44.4	5.6	5.6	64.6	27.7	7.7	36.9	64.9	27.7	7.4	37.2	-
ZENUWSTELSEL	8	75.0	25.0	0.0	50.0	62.0	33.9	4.0	28.1	59.5	36.4	4.1	23.1	+
PSYCHISCHE PROBLEMEN	7	85.7	14.3	0.0	71.4	63.1	29.3	21.9	33.7	63.4	28.9	22.0	34.4	+
TR.RESPIRATORIUS	15	73.3	20.0	6.7	53.3	77.2	17.3	12.1	59.8	78.3	16.9	11.5	61.5	0
HUD EN SUBCUTIS	10	90.0	10.0	0.0	80.0	76.6	18.5	24.9	58.1	77.5	18.2	24.3	59.2	+
ENDOCRIENE KLIJREN/METAB.	6	100.0	0.0	0.0	100.0	77.1	15.5	7.4	61.5	75.1	17.3	7.7	57.8	+
URINEWEGEN	6	66.7	33.3	0.0	33.3	67.4	28.4	4.1	39.0	66.8	29.0	4.2	37.7	0
ZWANGERSCHAP/BEVALL./ANTI	5	60.0	40.0	0.0	20.0	66.9	28.1	25.0	38.8	63.1	31.9	24.9	31.2	0
GESLACHTSORG./BORST VROUW	6	66.7	16.7	16.7	50.0	53.5	39.8	6.6	13.7	55.5	39.0	5.5	16.4	+
GESLACHTSORG./BORST MAN	5	60.0	40.0	0.0	20.0	69.3	22.7	28.0	46.6	68.9	24.9	26.2	43.9	-
HUISARTSGENEESK. THEORIE	17	64.7	23.5	11.8	41.2	69.8	17.8	30.0	52.0	70.5	18.5	28.7	51.9	-

TOTAALSCORE PROCENTUEEL STANDAARDDEVIATIE PROC.	146	70.5	22.6	6.8	47.9	68.8	23.5	7.8	45.3	68.9	24.0	7.0	44.9	
						7.4	6.1	6.3	12.0	6.9	5.9	6.1	11.3	
TOTAALSCORE ABSOLUUT STANDAARDDEVIATIE ABS.	146	103	33	10	70	100.4	34.2	11.3	66.2	100.6	35.1	10.3	65.6	
						10.8	8.9	9.2	17.5	10.1	8.6	8.9	16.6	

Toelichting -/o/+ : aangegeven wordt of men een lage, een gemiddelde of een hoge goed-minus-fout score heeft ten opzichte van de meest rechter kolom (Deeln. landel.).

Vervallen vragen: 5, 15, 39, 55, 56, 62, 81, 89, 101, 121, 123, 140,
151 en 153.

HUISARTSGENEESKUNDE, SV-IOH, UTRECHT
KENNISTOETS HUISARTSEM

Toetsdatum : FEBRUARI 1992
Toetsnaam : Kennistoets feb. 1992
Kandidaat nummer : Naam: Jaar:
Aantal geldige items : 146
Eigen groep : Huisartsen (niet-opleiders) (n=133)
Deeln. landel. : Alle landelijke deelnemende huisartsen (n=365)
UW RESULTAAT IS :

Toetsresultaten per categorie. Per individu en voor twee groepen (percentages)

CATEGORIE OMSCHRIJVING	AANT. VRAGEN	INDIVIDUEEL				EIGEN GROEP, n=133				DEELN. LANDEL., n=365				-/o/+
		GOED	FOUT	?	G-F	GOED	FOUT	?	G-F	GOED	FOUT	?	G-F	
< 15 jaar	35	74.3	17.1	8.6	57.1	76.0	17.5	12.1	58.5	76.4	17.7	11.6	58.7	0
15 tot 75 jaar	78	73.1	21.8	5.1	51.3	65.3	27.2	15.2	38.2	65.1	28.1	14.5	37.0	+
> 75 jaar	16	56.3	37.5	6.3	18.8	68.6	24.3	25.8	44.2	69.6	23.9	25.3	45.6	-
niet leeftijdsgebonden	17	64.7	23.5	11.8	41.2	69.8	17.8	30.0	52.0	70.5	18.5	28.7	51.9	-

Toetsresultaten per vak. Per individu en voor twee groepen (percentages)

VAK OMSCHRIJVING	AANT. VRAGEN	INDIVIDUEEL				EIGEN GROEP, n=133				DEELN. LANDEL., n=365				-/o/+
		GOED	FOUT	?	G-F	GOED	FOUT	?	G-F	GOED	FOUT	?	G-F	
diagnose	48	64.6	22.9	12.5	41.7	68.1	23.7	18.6	44.4	67.8	24.6	18.0	43.2	0
medicamenteus beleid	43	76.7	20.9	2.3	55.8	67.8	25.4	18.4	42.4	68.6	25.3	17.7	43.2	+
niet-medicamenteus beleid	31	74.2	25.8	0.0	48.4	69.8	24.5	8.9	45.3	69.4	25.4	8.4	44.0	0
epidemiologie	9	77.8	11.1	11.1	66.7	72.2	18.9	42.3	53.3	72.6	19.7	41.0	52.9	+
rest	15	60.0	26.7	13.3	33.3	69.5	17.5	12.9	52.0	70.4	18.2	11.5	52.2	-

TOTAALSCORE PROCENTUEEL STANDAARDDEVIATIE PROC.	146	70.5	22.6	6.8	47.9	68.8	23.5	7.8	45.3	68.9	24.0	7.0	44.9	
						7.4	6.1	6.3	12.0	6.9	5.9	6.1	11.3	
TOTAALSCORE ABSOLUUT STANDAARDDEVIATIE ABS.	146	103	33	10	70	100.4	34.2	11.3	66.2	100.6	35.1	10.3	65.6	
						10.8	8.9	9.2	17.5	10.1	8.6	8.9	16.6	

Toelichting -/o/+ : aangegeven wordt of men een lage, een gemiddelde of een hoge goed-minus-fout score heeft ten opzichte van de meest rechter kolom (Deeln. landel.).

Vervallen vragen: 5, 15, 39, 55, 56, 62, 81, 89, 101, 121, 123, 140,
151 en 153.

BIJLAGE 10 VOORBEELDPAGINA KENNISTOETS VAGINAAL BLOEDVERLIJES

Kennistoets Vaginaal Bloedverlies najaar 1992

*Vaginaal Bloedverlies***Casus 2**

Mevrouw De Lange, 45 jaar, heeft sinds een half jaar hevige, langdurende menstruaties die wel regelmatig zijn. Zij vloeit 7 dagen per maand, en dan zo heftig dat zij de eerste paar dagen de deur niet uit kan. 's Nachts moet zij luiers dragen en plastic broeken. Zij vloeit duidelijk méér en langer dan voorheen, vaak ook met stolsels. Pijn heeft zij daarbij niet. Momenteel vloeit zij niet. Zij wil voor haar klachten behandeld worden. Mevrouw De Lange is niet zwanger. Bij gynaecologisch onderzoek vindt de huisarts geen afwijkingen. Hij concludeert dat er sprake is van een hormonale regulatiestoornis* en schrijft mevrouw progestagenen voor van de 15e tot de 25e cyclusdag. Er bestaan geen contra-indicaties t.a.v. hormonale therapie.

3. In dit geval heeft behandeling met progestagenen de voorkeur boven behandeling met OAC. [onjuist]
4. Indien progestagenen voorgeschreven worden dient deze behandeling tenminste gedurende zes cycli te worden voortgezet. [onjuist]

Een paar maanden later neemt mevrouw opnieuw contact op. Haar klachten zijn niet verbeterd, maar het lukt haar ook slecht om de medicatie volgens voorschrift in te nemen. Zij wil eigenlijk ook liever geen hormonen. De huisarts schrijft nu NSAID's (Naproxen 2x 500 mg) voor.

5. Dit beleid is op dit moment correct. [juist]

* Hormonale regulatiestoornis: er is geen onderliggende afwijking aantoonbaar. De verstoring van het bloedingspatroon berust waarschijnlijk op een verstoring in de hormonale sturing van de opbouw van het endometrium.

BIJLAGE 11 INVENTARISATIE KENNISELEMENTEN NHG-STANDAARD CHOLESTEROL

S. Mol/M. Pollemans, december 1992

Epidemiologie

- met algemene GVO-programma's wordt een cholesteroldaling van 5% bereikt;
- van de volwassenen tot 60 jaar in Nederland heeft een vijfde een verhoogd cholesterolgehalte van > 6.5 mmol/l;
- slechts een derde van de volwassenen tot 60 jaar heeft een ideaal cholesterolgehalte van < 5.0 mmol/l;
- de voorspellende waarde van een verhoogd cholesterolgehalte is verschillend voor mannen en vrouwen;
- uit een aantal studies is gebleken dat verlaging van het cholesterolgehalte wel een daling in het aantal hartinfarcten geeft, maar niet leidt tot verlaging van de totale sterfte;

Diagnostiek

- case finding: bij wie moet men het cholesterolgehalte bepalen?
 - leeftijdsgrenzen (18 - 65 jaar);
 - xanthomen;
 - xanthelasmata (< 40 jaar);
 - troebel serum;
 - coronaire hartziekten bij patiënt zelf;
 - coronaire hartziekten bij eerstegraads familie < 60 jaar;
 - hypertensie;
 - diabetes mellitus;
 - familiale hyperlipidemie in de familie;
- roken als enige risicofactor is onvoldoende indicatie tot meting cholesterol;
- om een verhoogd cholesterolgehalte vast te stellen, zijn minimaal drie metingen nodig, met een periode van 1 tot 2 weken tussen twee metingen;
- een cholesterolbepaling hoeft niet nuchter;
- bepaling van HDL en triglyceriden dient nuchter te geschieden;
- HDL en triglyceriden worden pas gemeten als het cholesterolgehalte bij herhaling groter is dan 6.4 mmol/l, dieet niet helpt en overwogen wordt tot medicamenteuze therapie over te gaan;

Therapie

- bij een cholesterolgehalte van 5.1 - 6.4 mmol/l wordt een voedingsadvies gegeven;
- niet alleen cholesterol, maar met name de totale hoeveelheid verzadigd vet is van belang;
- betrek bij beslissingen over het beleid bij hypercholesterolemie ook de andere risicofactoren;
- bij verhoogde triglyceriden geen galzuurbindende harsen, maar fibraten voorschrijven;
- van de cholesterol-syntheseremmers is nog niet aangetoond dat ze de coronaire hartziekten-incidentie verlagen;
- cholesterol-syntheseremmers alleen voorschrijven als anticonceptie verzekerd is;
- de streefwaarde bij cholesterol-behandeling is 6.5 mmol/l;
- wanneer na een jaar behandelen met dieet en medicatie, het cholesterolgehalte hoger blijft dan 8.0 mmol/l dient verwezen te worden naar een specialist.

BIJLAGE 12 VOORBEELDPAGINA KENNISTOETS CHOLESTEROL

Kennistoets Cholesterol voorjaar 1993

Diagnostiek en Beleid

Maurice Dingemans, IC-verpleegkundige, 29 jaar, heeft op een beurs met medisch-technische snufjes zijn serumcholesterolgehalte laten bepalen. Dit bleek 6.3 mmol/l te zijn. Er zijn verder geen risicofactoren. Zijn huisarts raadt hem aan de bepaling te laten herhalen.

18. Volgens de NHG-Standaard Cholesterol verdient het inlassen van een wachttijd (van ongeveer twee weken) vóór de tweede bepaling, de voorkeur boven onmiddellijke herhaling van de bepaling. [JUIST]

De huisarts vertelt Maurice dat er in totaal drie bepalingen nodig zijn om een verhoogd serumcholesterolgehalte vast te stellen.

19. Als uitgangswaarde voor het uitstippelen van beleid wordt de hoogste waarde gebruikt. [ONJUIST]

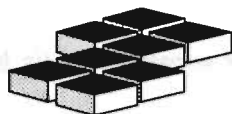
De (op de juiste wijze bepaalde) uitgangswaarde blijkt 6.3 mmol/l te zijn.

20. Bij een dergelijke waarde beveelt de NHG-Standaard Cholesterol een voedingsadvies aan. [JUIST]

BIJLAGE 13 SATISFACTIEVRAGENLIJST ONDERWERPGEBONDEN TOETSEN

Samenwerkingsverband Interfacultair Overleg Huisartsgeneeskunde

Uitvoerend Bureau



GRAAG INVULLEN ALS DE TWEEDE KENNISTOETS (DE NATOETS) IS BEANTWOORD.

Deelnemerscode:

--	--	--	--	--	--	--	--

SATISFACTIE-VRAGENLIJST KENNISTOETS VAGINAAL BLOEDVERLIES

	omcirkelen				niet te	
	geheel eens	eens	neutraal	oneens	geheel oneens	beoordelen/ geen mening
1. Ik vond het maken van de kennistoets, voorafgaand aan de nascholing, een zinvolle introductie op het programma.	1	2	3	4	5	6
2. De inhoud van de eerste kennistoets was een goede afspiegeling van de inhoud van het nascholingsprogramma.	1	2	3	4	5	6
3. De inhoud van de tweede kennistoets was een goede afspiegeling van de inhoud van het nascholingsprogramma.	1	2	3	4	5	6
4. Ik vond het lastig om 'Juist-Onjuist-Vraagteken'-vragen te beantwoorden.	1	2	3	4	5	6
5. Ik vond de kennistoetsen bruikbare hulpmiddelen om mijn eigen kennislacunes op het gebied van vaginaal bloedverlies vast te stellen.	1	2	3	4	5	6

6. Hoeveel tijd had u nodig voor het beantwoorden van de eerste toets? minuten

7. Hoeveel tijd had u nodig voor het beantwoorden van de tweede toets? minuten

De hierna volgende vragen hebben betrekking op uw oordeel over de toepasbaarheid (in het algemeen) van onderwerpgebonden kennistoetsen in de nascholing.

	omcirkelen				niet te	
	geheel eens	eens	neutraal	oneens	geheel oneens	beoordelen/ geen mening
Het maken van (korte) kennistoetsen voorafgaand aan nascholingsprogramma's, vind ik:						
8. nuttig als 'warming up' (d.w.z. als introductie op het programma);	1	2	3	4	5	6
9. nuttig als hulpmiddel om tot deelname aan een bepaald nascholingsprogramma te besluiten;	1	2	3	4	5	6
10. geen zinvolle activiteit.	1	2	3	4	5	6

OVERIGE OPMERKINGEN EN COMMENTAAR OP DE KENNISTOETSEN

Wat betreft de kennistoetsing in het algemeen:

Wat betreft de kennistoetsen vaginaal bloedverlies:

Wat betreft de vragen die in de kennistoetsen werden gesteld:

BIJLAGE 14 INSTRUCTIEBRIEF HUISARTSEN-DEELNEMERS ONDERWERPGEBONDEN TOETSING

Aan: huisartsen die deelnemen aan het nascholingsprogramma 'vaginaal bloedverlies'.

21 september 1992

US92250.MP

Informatie over de kennistoetsing rondom het nascholingsprogramma

Geachte huisarts,

Rondom het nascholingsprogramma 'vaginaal bloedverlies' waaraan u binnenkort deelneemt, wordt, in samenwerking met het Nederlands Huisarts Genootschap, een kennistoets-project georganiseerd. De bedoeling van dit project is na te gaan of het organiseren van kennistoetsen rondom nascholingsprogramma's, voor de deelnemers aan die programma's een zinvolle activiteit is. Het project wordt uitgevoerd door het landelijk Samenwerkingsverband van de acht beroepsopleidingen tot huisarts en de Werkgroep Onderzoek Kwaliteit huisartsgeneeskunde (WOK).

Om misverstanden te voorkomen: het gaat niet om het beoordelen van het individuele kennisniveau van de deelnemers. Het gaat er in het kennistoets-project om dat er na afloop van het project iets kan worden gezegd over de waarde van de methode van kennistoetsing in de nascholing. Daarbij gelden zaken als: hoe moet een goede kennistoets eruit zien, wat vinden de deelnemers ervan, waarvoor vinden de deelnemers kennistoetsen het meest geschikt, hoe ziet een adequate feedback eruit, is het mogelijk met behulp van deze toetsen effect van nascholingsprogramma's te bepalen, enz. De deelnemers krijgen na afloop van de toetsing (schriftelijk en individueel) een overzicht van de eigen resultaten, voorzien van de gemiddelde resultaten van hun groepsgenoten.

Wat is de bedoeling:

U vult, vlak voordat de nascholing plaatsvindt, een korte kennistoets in over het onderwerp vaginaal bloedverlies. Daartoe vindt u bij deze brief een kennistoets en een antwoordformulier.

U wordt verzocht dit antwoordformulier ingevuld mee te nemen naar de nascholings-bijeenkomst.

Naar verwachting kost het invullen van het antwoordformulier u een kwartier tot 20 minuten. In dit pakketje treft u tevens een korte vragenlijst naar enkele achtergrondkenmerken. Wij verzoeken u ook deze vragenlijst ingevuld mee te nemen naar de nascholingsbijeenkomst.

Na afloop van de nascholing zult u opnieuw worden gevraagd een korte kennistoets te beantwoorden, alsmede een lijstje met enkele stellingen waarop u kunt aangeven hoe u de gehele kennistoetsing heeft ervaren.

Enige tijd na afloop van het nascholingsprogramma kunt u een schriftelijk overzicht van uw toetsresultaten, voorzien van de gemiddelden van uw groepsgenoten, tegemoet zien.

Informatie over de kennistoetsen:

De kennistoetsen waarvan wij u vragen ze in te vullen, zijn inhoudelijk afgestemd op het nascholingsprogramma 'vaginaal bloedverlies' waarover door het NHG een standaard in voorbereiding is. De kennistoetsen weerspiegelen ook deze standaard. De toetsvragen hebben de volgende vorm: ze zijn geformuleerd als stellingen die meestal door een casusbeschrijving vooraf worden gegaan. Deze stellingen kunnen worden beantwoord met 'juist', 'onjuist' of met een 'vraagteken'. Het vraagteken kunt u invullen als u niet weet wat het antwoord is. Het vraagteken is dus niet bedoeld om aan te geven hoe zeker u van uw antwoord bent.

Het is voor de verwerking van de gegevens van groot belang dat u uw antwoordformulier met potlood invult. Het antwoordformulier is een doordrukformulier. De onderste bladzijde kunt u zelf houden; het bovenste formulier kunt u, na invulling, in de bijgevoegde retourenveloppe stoppen.

Procedure:

Voor het gemak hebben we alle informatie voor u nog eens op een rij gezet:

enkele dagen voor de nascholing:

- u ontvangt een pakket met:
 - een toelichtende brief (dwz de brief die u nu leest);
 - een toetsboekje (de blauwe voortoets);
 - een voorgecodeerd antwoordformulier;
 - een voorgecodeerde vragenlijst naar achtergrondinformatie;
 - een retourenveloppe.
- u wordt verzocht:
 - de vragenlijst naar achtergrondkenmerken in te vullen;
 - het antwoordformulier over de kennistoets in te vullen;
 - beide ingevulde formulieren in de retourenveloppe te stoppen en mee te nemen naar de nascholingsbijeenkomst.

op de nascholingsavond:

- u levert de retourenveloppe in bij uw WDH-contactpersoon;
- de nascholing vindt plaats;
- u ontvangt een tweede pakketje met:
 - een tweede toetsboekje (de groene natoets);
 - een voorgecodeerd antwoordformulier;
 - een satisfactievragenlijstje;
 - een retourenveloppe.
- u wordt verzocht:
 - het antwoordformulier over de kennistoets in te vullen;
 - de satisfactievragenlijst in te vullen;
 - beide formulieren in de retourenveloppe te stoppen en deze aan de WDH-contactpersoon te overhandigen.

na afloop van de nascholing:

- u ontvangt persoonlijk, schriftelijk feedback;
- u ontvangt een cadeaubon van f. 50,-.

Wellicht ten overvloede willen wij u er toch op wijzen dat de kennistoetsen, wil het onderzoek naar de toepasbaarheid ervan in de nascholing enige kans van slagen hebben, moeten worden ingevuld

zonder dat daarbij naslagwerken worden gebruikt en zonder dat daarbij onderling over de vragen wordt overlegd.

Wij hopen u hiermee van voldoende informatie te hebben voorzien. Mocht u op dit moment, of later, nog vragen hebben, dan kunt u zich altijd tot ons richten. Wij hopen dat u de toetsen met veel plezier maakt!

Hoogachtend, en met vriendelijke groeten:

namens ontwikkelaars en onderzoekers van het huisartsgeneeskundige kennistoetsproject,

A handwritten signature in black ink, appearing to be 'Marjan Poffemans', written over a horizontal line.

Marjan Poffemans, arts/onderwijskundige

Saskia Mol, huisarts

BIJLAGE 15 VOORBEELD INDIVIDUELE FEEDBACK ONDERWERPGEBONDEN TOETSING, PLUS TOELICHTENDE BRIEF

Aan : de huisartsen die hebben deelgenomen aan het nascholingsprogramma 'vaginaal bloedverlies'.

5 november 1992

US92339.MP

Feedback kennistoets 'vaginaal bloedverlies' najaar 1992

Geachte heer, mevrouw,

Hierbij ontvangt u uw persoonlijke feedback op de kennistoetsen die u heeft ingevuld voor en na deelname aan het nascholingsprogramma over vaginaal bloedverlies enkele weken geleden. De antwoordsleutels van beide toetsen sturen we u hierbij ook toe.

De feedback bestaat uit 1 formulier waarop u kunt zien wat uw persoonlijke resultaat was, wat het gemiddelde resultaat was van uw eigen groep, en tenslotte wat het gemiddelde resultaat was van alle huisartsen die in deze periode de betreffende toets hebben gemaakt.

In de kop van het formulier ziet u uw persoonlijke code (kandidaatnummer).

Zoals u in de tabel ziet zijn de scores berekend over 25 vragen. Eén vraag is na afloop uit de berekeningen verwijderd omdat de vraag een strikvraag bleek te zijn. Het betreft hier vraag 19 uit de voortoets. De parallelvraag uit de natoets (vraag 10) is bij de berekening van de resultaten over de natoets ook verwijderd.

De scores zijn eerst weergegeven als percentage vragen dat u van de toets goed, fout of met een vraagteken heeft beantwoord. Daaronder zijn ook de absolute scores weergegeven.

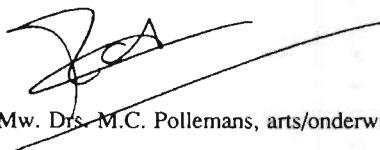
Tot zover onze toelichting op het feedbackformulier. Mocht u naar aanleiding van de toets, of de resultaten, nog vragen hebben, dan kunt u zich daarmee altijd tot ons richten.

Wij danken u voor uw bereidwilligheid de toetsvragen te beantwoorden. We hopen dat u de toetsen met plezier heeft ingevuld.

Mocht u, of de groep waartoe u behoort, belangstelling hebben om vaker aan landelijke, algemene, of meer specifieke, onderwerpgebonden kennistoetsing deel te nemen, dan is dat in principe mogelijk. Neemt u dan contact met ons op. Het adres en telefoonnummer van het Uitvoerend Bureau dat de kennistoetsing ontwikkelt, treft u aan op deze brief.

Wij hopen u hiermee van voldoende informatie te hebben voorzien. Namens onderzoekers en ontwikkelaars van de huisartsgeneeskundige kennistoetsen,

hoogachtend,



Mw. Drs. M.C. Pollemans, arts/onderwijskundige

Bijlagen: - persoonlijk feedbackformulier kennistoetsing 'vaginaal bloedverlies';
- antwoordsleutels.

(1 pagina voorbeeld feedback vag. bloedverlies)

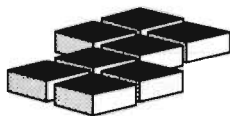
KANDIDAAT:
GROEP :

R E S U L T A T E N																
Omschrijving	Aantal Vragen	INDIVIDUEEL			UW GROEP (n=19)						ALLE DEELNEMERS (n=66)					
		Goed	Fout	?	Goed	std	Fout	std	?	std	Goed	std	Fout	std	?	std
Voortoets (procentueel)	25	72.0	20.0	8.0	64.4	13.3	30.1	11.9	4.4	5.0	60.9	12.8	30.0	10.8	8.2	8.2
Voortoets (absoluut)	25	18.0	5.0	2.0	16.1	3.3	7.5	3.0	1.1	1.3	15.2	3.2	7.5	2.7	2.0	2.1
Natoets (procentueel)	25	72.0	28.0	0.0	73.0	7.7	25.0	9.0	2.0	4.4	80.3	9.8	18.1	9.7	1.6	3.4
Natoets (absoluut)	25	18.0	7.0	0.0	18.3	1.9	6.2	2.2	0.5	1.1	20.1	2.4	4.5	2.4	0.4	0.8

Aantal vervallen vragen is 1

BIJLAGE 16 RETENTIEVRAGENLIJST KENNISTOETS OVER CHOLESTEROL

Samenwerkingsverband Interfacultair Overleg Huisartsgeneeskunde



Uitvoerend Bureau

GRAAG INVULLEN ALS DE KENNISTOETS IS BEANTWOORD.

Deelnemerscode:

--	--	--	--	--	--	--	--	--	--

VRAGENLIJST RETENTIE-KENNISTOETS CHOLESTEROL

		omcirkelen				niet te	
		geheel eens	eens	neutraal	oneens	geheel oneens	beoordelen/ geen mening
Deelname aan de nascholing over cholesterol heeft ertoe geleid dat ik:							
1.	- actiever ben geworden in het opsporen van hypercholesterolemie;	1	2	3	4	5	6
2.	- een ander beleid ben gaan voeren ten aanzien van patiënten met hypercholesterolemie.	1	2	3	4	5	6
3.	Mijn oordeel achteraf over het nascholingsprogramma over cholesterol is positief.	1	2	3	4	5	6
4.	Ik vond het lastig om 'Juist-Onjuist-Vraagteken'-vragen te beantwoorden.	1	2	3	4	5	6
5.	Hoeveel tijd had u nodig voor het beantwoorden van de toets? minuten					

6. Heeft u in de tijd die is verstreken
sinds de nascholing over cholesterol,
extra aandacht besteed aan het onder-
werp cholesterol (bv. via literatuur,
(na)scholing, anderszins)? ja / nee

Zo ja:

Kunt u toelichten in welke vorm?



OVERIGE OPMERKINGEN EN COMMENTAAR OP DE KENNISTOETS

Wat betreft de kennistoetsing in het algemeen:

Wat betreft de kennistoets cholesterol:

Wat betreft de vragen die in de kennistoets werden gesteld:

**BIJLAGE 17 VOORBEELDPAGINA KENNISTOETS CHOLESTEROL
(MEERKEUZEVRAGEN)**

Kennistoets Cholesterol voorjaar 1993

Diagnostiek en Beleid

Maurice Dingemans, IC-verpleegkundige, 29 jaar, heeft op een beurs met medisch-technische snuffjes zijn serumcholesterolgehalte laten bepalen. Dit bleek 6.3 mmol/l te zijn. Er zijn verder geen risicofactoren. Zijn huisarts raadt hem aan de bepaling te laten herhalen.

18. Wat betreft de aanbevolen wachttijd tussen twee bepalingen, meldt de NHG-Standaard Cholesterol het volgende:

- * a. hanteren van een termijn van 1 tot 2 weken,
- b. hanteren van een termijn van 2 tot 4 weken,
- c. hanteren van een termijn van 4 tot 8 weken,
- d. er bestaat geen reden om een wachttijd in te lassen.

De huisarts vertelt Maurice dat er in totaal drie bepalingen nodig zijn om een verhoogd serumcholesterolgehalte vast te stellen. Als uitgangswaarde voor het uitstippelen van beleid wordt volgens de NHG-Standaard Cholesterol genomen:

19. a. de laagste waarde,
b. de hoogste waarde,
c. het gemiddelde van de twee hoogste waarden,
* d. het gemiddelde van de drie waarden.

De (op de juiste wijze bepaalde) uitgangswaarde blijkt 6.3 mmol/l te zijn. Bij een dergelijke waarde beveelt de NHG-Standaard Cholesterol het volgende aan:

20. a. controle van het serumcholesterolgehalte na een half jaar,
b. controle van het serumcholesterolgehalte na een jaar,
* c. geven van een voedingsadvies,
d. instellen van medicamenteus beleid.

**BIJLAGE 18 VOORBEELD INDIVIDUELE FEEDBACK MEERKEUZETOETS,
PLUS TOELICHTENDE BRIEF**

Aan de huisartsen die hebben deelgenomen aan de multiple choicetoets 'cholesterol'

28 januari 1994

US94023.MP

Feedback kennistoets 'cholesterol' 1993

Geachte heer/mevrouw,

Hierbij ontvangt u uw persoonlijke feedback op de kennistoets over cholesterol, die u enkele maanden geleden heeft ingevuld.

De feedback bestaat uit 1 formulier, waarop u kunt zien wat uw persoonlijke resultaat is, wat het gemiddelde resultaat is van uw eigen groep, en tenslotte wat het gemiddelde resultaat is van alle huisartsen die in deze periode de betreffende toets hebben gemaakt.

In de kop van het formulier ziet u uw persoonlijke code (kandidaatnummer).

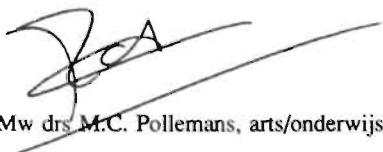
Wij danken u voor uw bereidwilligheid de toetsvragen te beantwoorden. Wij hopen dat u de toets met plezier heeft ingevuld.

Mocht u, of de groep waartoe u behoort, belangstelling hebben om vaker aan landelijke, algemene, of meer specifieke, onderwerpgebonden kennistoetsing deel te nemen, dan is dat in principe mogelijk. Neemt u dan contact met ons op. Het adres en telefoonnummer van het Uitvoerend Bureau dat de kennistoetsing ontwikkelt, treft u aan op deze brief.

Wij hopen u hiermee van voldoende informatie te hebben voorzien.

Namens de onderzoekers en ontwikkelaars van de huisartsgeneeskundige kennistoetsen,

hoogachtend,



Mw drs M.C. Pollemans, arts/onderwijskundige

Bijlage: persoonlijk feedbackformulier kennistoetsing 'cholesterol'

KANDIDAAT:
GROEP :

R E S U L T A T E N						
Omschrijving	Aantal Vragen	INDIVIDUEEL	UW GROEP (n=12)		ALLE DEELNEMERS (n=37)	
		Goed	Goed	stand.dev.	Goed	stand.dev.
MC-TOETS						
procentueel	30	43.3	45.6	12.9	49.2	12.8
absoluut	30	13.0	13.7	3.9	14.8	3.8

BIJLAGE 18 VOORBEELD INDIVIDUELE THERAPIE- en HET
PLAN VAN DE KLINISCHESCHIEDE

CURRICULUM VITAE

Marjan Pollemans werd geboren op 12 augustus 1952 te Breda. Daar behaalde zij in 1970 het diploma Gymnasium B aan de Scholengemeenschap Mencia de Mendoza. Vervolgens studeerde zij geneeskunde aan de Katholieke Universiteit in Nijmegen, waar zij in 1975 het doctoraal-examen en in 1977 het artsexamen behaalde. Daarna begon zij aan de Universiteit van Amsterdam een studie pedagogische wetenschappen, waarvoor in 1987 het doctoraal-examen werd afgelegd (hoofdvak algemene en vergelijkende onderwijskunde, bijvakken sociologie en bestuurskunde).

Van 1978 tot 1980 werkte zij bij het Centraal Instituut voor Toetsontwikkeling te Arnhem en van 1980 tot 1982 bij het Instituut voor Maatschappelijke Gezondheidszorg van de Erasmusuniversiteit te Rotterdam. Daarnaast was zij van 1979 tot 1985 verbonden aan de Rijksuniversiteit Limburg te Maastricht, waar zij deel uitmaakte van het Project Evaluatie van Studieresultaten van de Faculteit der Geneeskunde. Van 1985 tot 1987 was zij, namens het Nederlands Huisartsen Genootschap, werkzaam bij de vakgroep Huisartsgeneeskunde van de Rijksuniversiteit Utrecht. Zij was lid van de projectgroep Curriculumconstructie voor de (meerjarige) Beroepsopleiding tot Huisarts (CCBOH). Sinds 1987 is zij verbonden aan de Rijksuniversiteit Utrecht en werkzaam bij het Samenwerkingsverband van de Universitaire Huisartsopleidingen. Het Samenwerkingsverband heeft het ontwikkelen van een landelijk systeem van toetsing en evaluatie van de huisartsopleiding tot taak.

Zij is getrouwd en heeft drie kinderen.

ISBN 90-5278-159-1



9 789052 781594

datawyse | universitaire pers maastricht